DISCUSSION PAPER PI-1106

Longevity Hedge Effectiveness: A Decomposition

Andrew J.G. Cairns, Kevin Dowd, David Blake, and Guy D. Coughlan

August 2012

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University London
106 Bunhill Row
London EC1Y 8TZ
UNITED KINGDOM

http://www.pensions-institute.org/

Longevity Hedge Effectiveness: A Decomposition

Andrew J.G. Cairns, Kevin Dowd, David Blake, and Guy D. Coughlan[1]

First version: September 2010
This version: August 9, 2012

Abstract

We use a case study of a pension plan wishing to hedge the longevity risk in its pension liabilities at a future date. The plan has the choice of using either a customised hedge or an index hedge, with the degree of hedge effectiveness being closely related to the correlation between the value of the hedge and the value of the pension liability. The key contribution of this paper is to show how correlation and, therefore, hedge effectiveness can be broken down into contributions from a number of distinct types of risk factor. Our decomposition of the correlation indicates that population basis risk has a significant influence on the correlation. But recalibration risk as well as the length of the recalibration window are also important, as is cohort effect uncertainty. Having accounted for recalibration risk, additional parameter uncertainty has only a marginal impact on hedge effectiveness. Finally, the inclusion of Poisson risk only starts to become significant when the smaller population falls below about 10,000 members over age 50.

Our case study shows that, at least for medium and large pension plans, longevity risk can be substantially hedged using index hedges as an alternative to customised longevity hedges. As a consequence, when the hedger's population involves more than about 10,000 members over age 50, index longevity hedges (in conjunction with the other components of an ALM strategy) can provide an effective and lower cost alternative to both a full buy-out of pension liabilities or even to a strategy using customised longevity hedges.

Keywords: hedge effectiveness, correlation, mark-to-model, valuation model, simulation, value hedging, longevity risk, stochastic mortality, population basis risk, recalibration risk.

[1]Andrew J.G. Cairns: Maxwell Institute for Mathematical Sciences, and Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK. E: A.J.G.Cairns@hw.ac.uk. Kevin Dowd and David Blake: Pensions Institute, Cass Business School, City University, 106 Bunhill Row, London, EC1Y 8TZ, UK. Guy D. Coughlan: Pacific Global Advisers, 535 Madison Avenue, New York, NY 1022-4214, USA.

1 Introduction

Hedging the longevity risk in pension plans – the risk that, in aggregate, plan members live longer than anticipated – is becoming increasingly important. As more defined benefit pension plans close to future accrual and pension liabilities accordingly become crystallised, plan sponsors face the choice of selling their legacy pension liabilities or retaining them on their books and managing them.

The UK was the first country in the world to witness the development of both a buy-out market for pension liabilities and a longevity swap market to help sponsors hedge longevity risk as part of an asset-liability management (ALM) strategy.

With a buy-out, an insurance company, in exchange for a buy-out fee, takes over the plan liabilities and assets and takes on the responsibility for making the pension payments until the last plan member dies. A buy-out is known as an insurance indemnification solution, since all risks in the pension plan – the key ones being interest-rate, inflation-rate and longevity risk – are fully transferred from the sponsor to the insurer. The cost of a buy-out is high since the insurer has to post substantial regulatory capital to ensure that the pension payments will be made with a high degree of probability, as well as to ensure, *ex ante*, that the purchase price offers an adequate expected return relative to the risks being transferred. In addition to transferring all the pension assets, the sponsor might also need to make a cash payment to the insurer if the plan is in deficit, in order to fund the buy-out. Further, the sponsor foregoes the opportunity to manage the pension assets efficiently itself and so reduce the ultimate cost of the liability.

In contrast, a sponsor might decide to retain the pension plan and implement an ALM strategy, which broadly replicates the same economic effect as a buyout. This brings certain cost advantages. First, the sponsor saves making the buyout risk premium which would otherwise be paid to the insurer as compensation for taking on the risks associated with the pension plan. Second, the cost of each component of the ALM strategy can be separately negotiated and implemented, providing greater transparency, minimal upfront hedging costs (since the principal hedging instruments, interest-rate, inflation-rate and longevity swaps, have a zero value at execution) and flexibility in the timing and structure of implementation. However, the key disadvantage of such an ALM strategy – which has been dubbed a "do-it-yourself" (DIY) buy-out – is that the risks are not perfectly hedged. This is due to the idiosyncracies of each pension plan's membership and benefit structure. Swaps can hedge a significant proportion of the relevant risks in a given pension plan, but inevitably there will be some residual basis risk which cannot be hedged cost-effectively using capital market instruments.

This paper deals with the hedging of longevity risk, and so we will focus our remarks on this issue specifically. An ALM strategy might include the use of longevity swaps. However, there are different types of longevity swap and, accordingly, different levels of basis risk. A customised longevity swap takes into account the particular characteristics of each pension plan's demographics and benefit structure and is designed to maximise hedge effectiveness. An important alternative to this is an index swap that is linked to a relevant longevity index, rather than to the longevity experience of the pension plan members. As an example, the index might be related to the national population of the country in which the pension plan is domiciled (which we denote population 1 below). Since the composition of a pension plan's membership (which we denote population 2 below) will differ from that of the index, the hedge will inevitably involve greater basis risk (and hence lower effectiveness) than a customised swap. As a standardised product, an index swap has the advantage of being cheaper, less complex, and much easier to unwind. However, it only attempts to reduce longevity risk, rather than eliminate it completely.

Plan sponsors therefore face two key trade-offs. One is between the high costs and complete indemnification of a buy-out versus the lower costs and basis risk associated with a DIY-buyout/ALM strategy. The other, within the context of ALM, is between the higher costs and minimal basis risk of a customised longevity swap versus the lower costs and greater liquidity but higher basis risk associated with an index swap.

1.1 Analysis and evaluation of longevity hedges

In this paper, we examine the trade-off between customised and index longevity hedges. Coughlan et al. (2011) proposed a clear framework for "(i) developing an informed understanding of the basis risk, (ii) appropriately calibrating the hedging instrument and (iii) evaluating hedge effectiveness". In this paper, we follow closely Coughlan et al. (2011) both in terms of the framework and their main case study. However, the key difference, and the main contribution, in the present work is that whereas Coughlan et al. used a largely model-free bootstrapping approach to the evaluation of hedge effectiveness in their case study, we use a model-based simulation approach. As will be demonstrated later, this allows us to break down basis risk and the evaluation of hedge effectiveness into a number of components by switching on and off a number of key risk factors.

Our case study involves the use of England & Wales male mortality (the LifeMetrics index) to hedge liabilities linked to Continuous Mortality Investigation (CMI) male assured lives mortality. [2] The case study considers a value hedge (as opposed to a cashflow hedge) set up at time 0 of a pension plan liability's exposure to longevity risk at a single future valuation date, T.[3] The hedging instrument that we

[2]http://www.actuaries.org.uk/research-and-resources/pages/continuous-mortality-investigation/

[3]A value hedge seeks to hedge the present value of a sequence of future pension cash flows at a single future date, T. This contrasts with a cashflow hedge which consists of an asset strategy which delivers a sequence of cashflows that is as close as possible to the sequence of pension plan liability cashflows. Value hedging is fundamentally different from cashflow hedging. An effective value hedge can be achieved using a variety of hedging instruments, each of which can be quite different in style from the liability value being hedged. In contrast, a cashflow hedge generally requires a hedging instrument that is very similar in structure to the liability cashflows. Nevertheless, the

use will be a "cash-settled" deferred longevity swap (defined later). Decomposing the *correlation* between the hedging instrument and the liability values is broadly equivalent to decomposing the *effectiveness* of the hedge.

There are three key categories of factor that contribute to an assessment of hedge effectiveness or the correlation between a hedging instrument and the liability being hedged:

1. Factors related to population differences, including:
   - Population basis risk: this arises as a result of using a hedging instrument linked to a different reference population from that of the hedging population.
   - Mismatched cohorts especially at younger ages: typically, the hedger of population 2 will wish to hedge the longevity risk for an existing group of plan members with accrued pension rights: that is, there will be some historical data for that cohort. However, the hedger might choose to link the hedging instrument to a cohort born in a different year (resulting in an age mismatch). In theory, this reference cohort might be one for which there will be no data available until after time 0. In this case, the value of the hedging instrument at T has the cohort effect as an additional source of uncertainty that will have a detrimental impact on hedge effectiveness.[4]
2. Factors related to the model used for simulation, including:
   - The choice of model to be fitted to historical mortality data and how the parameters and latent state variables of this model will be calibrated.[5] This model will be used to simulate future mortality scenarios which will then, one by one, be fed into the valuation model discussed below.[6]
   - Parameter uncertainty: arises because the true values of the parameters of the simulation model used to generate future mortality scenarios

ideas that we present here can be easily adapted for other choices of hedging instrument. Value hedging is important in a number of circumstances, including: situations where meeting shorter-term solvency requirements is more onerous than meeting cashflows in the long term (e.g., Nielsen, 2010); mergers and acquisitions where pension plan value is significant relative to the operating business; hedging longevity risk associated with deferred pensions and annuities.

[4]This means that, in practice, linkage to a future cohort would be suboptimal and hence not to be recommended.

[5] In line with other authors in the longevity risk management space we use the term *calibration* rather than estimation. Krugman (2011) defines calibration as "tweaking the parameters of your model until it fits some aspects of the data, rather than flat-out estimating the model". Thus, *calibration* includes 'flat-out' estimation (which might be the case in our present context), but equally it includes cases where expert judgement in some form is employed, and this is often the case in practice: for example, in the setting of mortality improvement rates. We note that one study in a different field has coined the term *estibration* to describe this procedure (Balistreri and Hillberry (2005)).

[6]There is model risk associated with the simulation model, since we do not know the true model generating future mortality rates: we disregard this risk in this study.

and quantify longevity risk are unknown – this covers both the process parameters (i.e., parameters governing the dynamics of the underlying stochastic processes) and the latent state variables of the model (i.e., the underlying age, period and cohort effects).

- Poisson risk:[7] otherwise known as small-population risk or sampling variation; the risk that the mortality experience of a small group of people will differ from the underlying true mortality rate; the financial consequences can be magnified if there is significant variation between individuals in pension entitlements.[8]

3. Factors related to the model used for valuation at the future valuation date, T:

   - The choice of model to be used to value liabilities at time T. This model is likely to be different from the simulation model.[9]
   - Recalibration risk: the uncertainty in both future liability values and hedging-instrument values associated with the calibration and recalibration of the parameters of the valuation model used to project mortality beyond the valuation date, T. The valuation model contains a number of process parameters that are assumed to be remain constant over time. However, the model will normally be calibrated using the latest available data. Thus, the calibration will be dependent on the specific scenario under consideration, and will be based solely on observed deaths and exposures rather than assuming knowledge of the underlying latent state variables. The extent to which valuation model parameters vary from one simulation scenario to the next results in additional randomness in liability and hedging-instrument values at T. Recalibration risk is, therefore, heavily dependent on the scenarios generated by the simulation model and includes the influence of both parameter uncertainty and Poisson risk.
   - Recalibration window: the length of the lookback window over which the valuation model is estimated and subsequently recalibrated; this reflects a tradeoff between using more years of data to get a better estimate of the volatility in the data and using fewer years of data to get a better estimate of the current trend in mortality improvements; it has a direct influence on recalibration risk.

4. Factors related to the structure of the hedge, such as:

[7] So-called because deaths in the pension plan are generally assumed to follow a (conditional) Poisson distribution; see, e.g., Dahl et al. (2008), Li and Hardy (2011) and Biffis et al. (2010). See, however, Li et al. (2009) for alternative assumptions.

[8] We do not consider this so-called concentration risk explicitly in the present paper. However, we can note that this type of concentration of risk is more important in cashflow hedging problems.

[9] There is also model risk in respect of the valuation model; again we disregard this risk in this study.

- Choice of hedging instrument.
- Choice of maturity date, reference population and reference age(s).
- Sub-optimal or inaccurate hedge ratio.
- Robustness of the hedge ratio: the challenge is to devise strategies that can maximise hedge effectiveness and to find solutions that are robust relative to, for example, errors in the specification of the model and parameters, etc.
- Index versus customised hedges.
- Static versus dynamic hedges.[10]
- Multi-instrument[11] versus single-instrument hedges.

The above list is quite extensive and it would not be feasible to examine all possible factors in a single study. Nevertheless, ours is the first study to carry out a forensic analysis of what we anticipate being the most important risk factors in a longevity hedging context, namely population basis risk, cohort effect uncertainty, recalibration risk, the impact of the length of the recalibration window, parameter uncertainty, and Poisson risk.

Previous studies which have examined a smaller subset of risk factors include: Dahl et al. (2008, 2009), Plat (2009), and Coughlan et al. (2011). Earlier studies which have examined different hedging instruments, such as longevity swaps, deferred longevity swaps and other longevity-linked bond and derivative structures, include Blake and Burrows (2001), Blake et al. (2006), Coughlan et al. (2007), Loeys et al. (2007), Cairns et al. (2008), Coughlan (2009), Wills and Sherris (2010), and Blake et al. (2010). Previous studies which have looked at the value-hedging paradigm include Coughlan et al. (2011) – in terms of effective risk reduction when future cashflows are highly unpredictable – and Nielsen (2010) and Olivieri and Pitacco (2009) in the context of Solvency II.

We find in this paper that recalibration risk has an important role to play in the assessment of hedge effectiveness. This is because we have a limited amount of historical data, leading to parameter uncertainty in both process parameters and the underlying state variables. This paper is the first to consider recalibration risk in the longevity literature. However, the concept is familiar elsewhere in the finance literature. The key issue is that model parameters that are assumed to remain constant are, in fact, recalibrated on a regular basis: partly because of parameter uncertainty and partly because the "true" model generating prices is different from the model being calibrated against these prices (e.g., the Black-Scholes model). The result is a sequence of calibrations that is inconsistent with the constant-parameter assumption. The fact that, for example, equity volatility is known to vary over time

[10] In this paper, we only consider static hedges. However, especially if there were a liquid market in appropriate hedging instruments, the hedge ratios could be modified from time to time between commencement of the hedge (time 0) and the target valuation date (time T).

[11] For example, the use of two or more deferred longevity swaps with different reference ages.

(as well as over strike prices and maturity dates) rather than remain constant, results in derivatives desks having to hedge against changes in volatility (vega hedging).

A related, but different, form of calibration risk concerns the method use to calibrate a complex model to a given set of market data (see, for example, Detlefsen and Härdle, 2007). The nearest equivalent in the mortality modelling context would, perhaps, be the choice between the conditional Poisson model for death counts and some other distribution (e.g., the normal distribution assumed by Lee and Carter, 1992).

We also find that the major determinants of correlation, and therefore hedge effectiveness, are population basis risk and the length of the recalibration window. Lesser, but still important factors are: parameter uncertainty (other than recalibration risk) and the reference age for the hedging instrument (especially if the reference age is at the lower end of the age range analysed).

1.2 Structure of the paper

The remainder of the paper is organised as follows. Section 2 sets out a case study of a pension plan that is considering hedging the longevity risk it faces using either a customised or an index longevity hedge. Section 3 outlines the five steps in constructing and evaluating the hedge using the very general framework of Coughlan et al. (2011) and discusses the role of correlation (between the values of the hedging instrument and the liability) in determining the level of hedge effectiveness. Section 4 describes the data and stochastic mortality model that we will use. Section 5 discusses how the model is used for both (i) simulating future mortality rates and (ii) valuing both the liability (a type of deferred annuity) and the hedging instrument. Although the choice of simulation model is independent of the choice of valuation model, we use the same model for convenience. Section 6 is the key numerical section that focuses on the correlation between the value of the pension liability in our case study and the values of both customised and index-based hedging instruments and quantifies how the different risk factors influence these correlations. Finally, Section 7 concludes.

2 A case study: A customised versus index hedge

Our discussion is centred on a stylised case study involving a UK pension plan consisting of male members only, which pays no spouses' or dependants' benefits. We evaluate hedging instruments that hedge the longevity risk associated with the value of the pension liability. The pension plan members will be assumed to have underlying mortality rates that are the same as the CMI male assured lives dataset and the pension liability will be calculated with reference to current and projected CMI mortality. This choice is because the CMI population has a very different mortality profile from the national population (see for example, Coughlan et al., 2011), thereby allowing us to easily incorporate population basis risk into the discussion. In order to hedge the longevity risk in the pension plan, we will consider both a hedging instrument linked to CMI male mortality (in the case of a customised hedge) and one linked to England & Wales (EW) male mortality (in the case of an index hedge). At the time we conducted this study, data were available for both populations up to the end of 2005 (time $t = 0$).[12]

Now define $a_k(T, x)$ as the value at T of a pension (or, equivalently, a life annuity) of £1 per annum payable annually in arrears from time T until death to a male aged x at time T in population k:

$k = 1$	England & Wales, males
$k = 2$	CMI assured lives, males.

Interest rates will be assumed to be constant and equal to r per annum in order that we can focus attention on longevity risk. With this in mind we can represent the value of the pension as

$$a_k(T, x) = \sum_{s=1}^{\infty} (1 + r)^{-s} p_k^{fwd}(T, s, x) \tag{1}$$

where the forward (prospective) survival probability, $p_k^{fwd}(T, s, x)$, represents the best estimate at T, that an individual aged x at time T in population k will survive for a further s years. The forward survival probabilities will be evaluated using a specified valuation model that will be discussed in detail in Section 5.

Our objective is to hedge the longevity risk in the value of a pension liability $L(T) = a_2(T, x)$, where $T = 10$ years (i.e., the end of 2015) and $x = 65$.[13] In our case study, the chosen hedging instrument will be a "cash-settled" deferred longevity swap that exchanges, at time T, the present value of a series of fixed cashflows for the present value of a set of floating cashflows occurring after time T. The floating cashflows will be equal to the proportions of a cohort aged y in population k at time T that are still alive at times $T+1, T+2, \ldots$, while the fixed cashflows of $K(T+s)$ for $s = 1, 2, \ldots$ are fixed at time 0. Thus, the value at time T of the floating leg of the swap will be $a_k(T, y)$ (i.e., the same as the value of an annuity) and we will denote the value at T of the fixed leg by $\hat{a}_k^{fxd}(0, T, y)$, where the additional argument of 0 refers to the date on which the fixed leg was contracted (in other words, this denotes a deferred annuity agreed at time 0 with the first payment at time T). We will use $H(T)$ to denote the cash-settled value at T of the deferred longevity swap. In summary, we, therefore, have the following values at T based on information available at time T:

$$\begin{aligned} L(T) &= a_2(T, 65) \quad \text{where } T = 10 && (2) \\ \text{and} \quad H(T) &= a_k(T, y) - a_k^{fxd}(0, T, y). && (3) \end{aligned}$$

[12] In fact, EW data were available up to 2009. This potential mismatch is discussed in further detail in Cairns (2011b).

[13] Here $L(T)$ represents the discounted value at T of the future unknown liability cashflows at $T+1, T+2, \ldots$, and takes account of the information that is known at time T but not after time T.

3 Constructing and evaluating a hedge

3.1 The hedge effectiveness framework

Following the framework of Coughlan et al. (2004, 2011), there are five steps in constructing and evaluating a hedge – whether customised or index. These steps have been slightly modified to suit our case study and are outlined in Tables 1 to 3.

Step	**Case study details**
Step 1: Objectives	
Risk to be hedged	Liability value, $L(T) = a_2(T, x)$
Horizon	$T = 10$
Full or partial risk mitigation?	Partial risk reduction
Step 2: Hedging instrument	
Choice of instrument	Deferred longevity swap, value at T: $H(T) = a_k(T, x) - \hat{a}_k^{fxd}(0, T, x)$ (no collateral or margin calls)
Hedged position: static or dynamic?	Static: $P(h) = L(T) + h \times H(T)$
Step 3: Method for assessment of hedge effectiveness	
Risk metric	$Var\left(P(h)\right)$
Basis for hedge effectiveness	$1 - Var\left(P(h)\right)/Var\left(L(T)\right)$
Scenario generator	Two-population Age-Period-Cohort stochastic simulation model
Valuation model	2× One-population APC models with consistent projections
Step 4: Hedge effectiveness calculation	
Simulate future mortality rates up to T	See Table 2
Evaluate position at T	See Table 3
Calibrate hedge ratio	$h^* = -\rho_{LH} \times SD(L(T))/SD(H(T))$
Evaluate hedge effectiveness	(h^* minimises $Var(P(h))$)
Step 5: Detailed analysis and interpretation of results	

Table 1: Five steps in constructing and evaluating a hedge (adapted from Coughlan et al., 2011).

Step 1 in Table 1 requires a clear definition of the hedging objectives. This includes defining the position to be hedged and the hedge horizon, T. In our case study, the metric, or quantity at risk, to be hedged is the value of the liability, $a_2(10, 65)$, over a horizon of 10 years. This step also involves a clear definition of the risk to be hedged and whether to mitigate it entirely (indemnification) or whether to mitigate it partially (leaving some degree or other of residual basis risk).

	Population $k = 1$	**Population** $k = 2$
1	Past mortality rates for index population (up to time "$t = 0$")	Past mortality rates for pension plan (up to time "$t = 0$")
2	Fit two-population model	
3	Simulation of two-population underlying mortality rates for $t = 1, \ldots, T$	
4	Index population: Add Poisson risk to death counts	Pension plan: Add Poisson risk to death counts
5	Future scenarios for index mortality experience, $t = 1, \ldots, T$	Future scenarios for pension plan mortality experience, $t = 1, \ldots, T$

Table 2: Five stages of simulation

	Population $k = 1$	**Population** $k = 2$
Historical data + simulation results		
1A	Past mortality rates for index	Past mortality rates for pension plan
1B	+ Future mortality scenarios for index	+ Future mortality scenarios for pension plan
Valuation model		
2	Scenario + Model $\Rightarrow$ calibration for hedging instrument valuation	Scenario + Model $\Rightarrow$ calibration for pension plan liability valuation
3	Consistent valuation model mortality projections	
4	For each scenario: Index hedging instrument valuation	For each scenario: Pension plan liability valuation
5	Calculate hedge effectiveness	

Table 3: Five stages of evaluation

In step 2, we choose the hedging instrument, or instruments, that we will use to reduce the liability risk. In the present case, it will be a deferred longevity swap, with a choice of reference population, k,[14] maturity dates, T, and reference starting ages, y. The hedge will be a static value hedge.[15]

Step 3 is the crucial step of defining the method for hedge effectiveness assessment. This is important because an inappropriate choice can easily lead to misleading hedge effectiveness results. This step involves not only the risk metric used to assess hedge effectiveness but also the method in which it is applied. For our case study, we choose the variance in the value of the pension liability as the risk measure (the same as, for example, Li and Hardy, 2011). Hedge effectiveness then provides us with a proportionate assessment of how much the variance of the liabilty will fall as a result of hedging.

We take a prospective approach to hedge effectiveness assessment using forward looking simulation of future mortality rates (see Coughlan et al., 2004, for a discussion of this and other choices). The risk measure is derived from a large number of independent scenarios for mortality rates between time $t = 0$ and time T that are generated using a stochastic simulation model.[16]

There are two key stages in Step 4: simulation and valuation. First, there is a simulation stage that takes us from the present time $t = 0$ to time T (see Table 2). This requires a two-population stochastic mortality model[17] to be calibrated to historical data up to time $t = 0$ that can then be used to simulate future mortality rates for both populations to time T. Second, for each stochastic scenario up to time T, we need to be able to value the liability and hedging instrument at time T. Valuation of these requires us to project, at T, the future liability cash flows beyond time T (see Table 3). We, therefore, extend each sample path of mortality rates up to time T into a two-dimensional mortality table that projects beyond time T. The final year of the simulated scenario at time T gives us the base, one-dimensional mortality table, and the pattern of mortality improvements up to time T are used to turn this base mortality table into a two-dimensional set of projected mortality and survival rates that can be used to calculate annuity values at T. We are then in a position to evaluate hedge effectiveness.

In other words, the outcome from the simulation and valuation procedures is a bivariate distribution for the liability and hedging instrument values at T. This, in combination with our chosen measure of hedge effectiveness, allows us to calculate the optimal hedge ratio, h^*.

Step 5 analyses the results of steps 1 to 4. This includes testing the robustness of our

[14] $k = 1$ for an index swap and $k = 2$ for a customised swap.

[15] Dynamic hedging is not feasible except at potentially significant cost. Additionally, with our particular choice of liability and hedging instrument, dynamic hedging does not, in fact, result in a significantly better hedge.

[16] There are other methods of generating these scenarios, for example, Coughlan et al. (2011) used bootstrapping of historical data.

[17] This jointly models two related populations by recognising the interdependence between them.

solutions to the assumptions used in the calculations, as well as assessing whether the results make intuitive sense.

3.2 Correlation and hedge effectiveness

Ultimately, our aim is to measure the effectiveness of any hedging strategy that we might choose to adopt. Here we focus on a simple value-hedging setting where we consider a static hedge using a single hedging instrument.

Suppose that we have a future random liability with value $L = L(T)$ at time T. Alongside this, we have a hedging instrument that has value $H = H(T)$ at time T. Our hedged portfolio consists of the liability plus h units (the hedge ratio) of H and its value at T is $P(h) = L + h.H$.

If we use variance as our measure of risk,[18] hedge effectiveness is defined as $R^2(h) = 1 - Var[P(h)]/Var[L]$: that is, it measures the proportionate reduction in risk due to the hedge. The optimal hedge ratio per unit of liability, L, then becomes

$$h^* \quad = \quad -\rho\frac{SD(L)}{SD(H)} = -\frac{Cov(L,H)}{Var(H)}, \tag{4}$$

where $\rho = Cor(L,H)$ (see, for example, Coughlan et al., 2004, for a general discussion of the optimal hedge ratio in a hedge effectiveness context). We then have

$$R^2(h^*) = \rho^2 \ \text{ and } \ R^2(h) \quad = \quad \rho^2\left(1 - \frac{(h-h^*)^2}{h^{*2}}\right). \tag{5}$$

We can conclude from (5) that, in this simple situation with a static hedge and a single hedging instrument, it is sufficient for us to analyse the correlation between L and H. When comparing hedging instruments, the one that has the highest (absolute) correlation will deliver the highest optimal hedge effectiveness, provided the optimal hedge ratio is employed.

The use of variance (or standard deviation) as the risk measure is consistent with the objective of UK pension plans to eliminate as much risk as possible over a period of years.

4 Data and model

We will use EW and CMI data covering ages 50 to 89 and calendar years 1961 to 2005 (with 2005 treated as $t = 0$). The full range of these data is used to fit the two-population stochastic mortality model specified below. This model plus parameter estimates – with some, but not all, experiments incorporating parameter uncertainty – is then used to simulate mortality rates at ages 50 to 89 for the years 2006 to 2015.

[18] For a short discussion of variance as our risk metric and some alternatives, see Appendix A.

The choice of age range means that the CMI cohort aged 65 in 2015 – the cohort that we refer to in our liability $L(T) = a_2(T, 65)$ – was aged 55 in 2005. Thus, our initial dataset up to 2005 already provides us with an estimate of the cohort effect that will be used in the evaluation of $a_2(T, 65)$.

For valuation purposes, actuaries will be assumed to have data available from 1961 up to the end of 2015. However, a projection model intended to project beyond time T will only be calibrated using data from the most recent 20 years (1995 to 2015) in order to capture the most recent trend in mortality rates. The assumption of a 20-year lookback window is consistent with market practice, although not all practitioners, of course, will use exactly 20 years (see the discussion in Dowd et al., 2010b).[19]

We will use the two-population Age-Period-Cohort (APC) model for $m_k(t, x)$, the population-k death rate, discussed in Cairns et al. (2011b).[20] Specifically, we assume that

$$\log m_k(t, x) = \beta^{(k)}(x) + \frac{1}{n_a}\kappa^{(k)}(t) + \frac{1}{n_a}\gamma^{(k)}(t - x) \tag{6}$$

where: t is the calendar year; x is the age last birthday; n_a is the number of individual ages covered by the dataset;[21] $\beta^{(1)}(x)$ and $\beta^{(2)}(x)$ are the population 1 and 2 age effects, respectively; $\kappa^{(1)}(t)$ and $\kappa^{(2)}(t)$ are the corresponding period effects; $\gamma^{(1)}(c)$ and $\gamma^{(2)}(c)$ are the corresponding cohort effects; and $c = t - x$ = cohort year of birth. Given $m_k(t, x)$, actual death counts, $D_k(t, x)$, for population k and age x in year t are assumed to have a conditional Poisson distribution with mean $m_k(t, x)E_k(t, x)$, where $E_k(t, x)$ is the central exposed to risk, both in historical model fitting and in the forecasts. See Section 5.1 for further details.

This model is one of the simplest that incorporates both random period and cohort effects. Our reasons for including a cohort effect are twofold. First, cohort effects have been found to be significant in a number of countries (e.g., England & Wales, France, Germany, Japan and Italy; see Cairns et al., 2011a). Second, when we consider possible hedges of longevity risk, we build on the observations of Cairns et al. (2011b) to demonstrate that the presence of a significant cohort effect can have a material impact on correlation and, implicitly, hedge effectiveness in a way that would not be evident if a model with no stochastic cohort effect were used.

The stochastic elements in our model (i.e., the period and cohort effects) are structured in a way that assumes that one population is large and the other population is a small (sub-)population. Thus (see Cairns et al., 2011b, for further discussion),

- Large population 1

[19] However, even if our use of the Age-Period-Cohort model for valuation is correct, we cannot be sure what length of lookback window, W, valuers will use. Indeed, W, might be considered to be a source of *Knightian uncertainty*: W is not just uncertain, but the degree of uncertainty is not quantifiable. Dealing with W as a source of uncertainty is discussed further by Cairns (2011b).

[20] Alternative multi-population models have been proposed by Li and Lee (2005), Dahl et al. (2008, 2009), Jarner and Kryger (2011), Plat (2009) and Dowd et al. (2011a).

[21] For example, our dataset covers ages 50 to 89, so $n_a = 40$.

- $\kappa^{(1)}(t)$ is modelled as a random walk with drift ν_1.
- $\gamma^{(1)}(c)$ is modelled as an AR(2) process mean-reverting to a linear trend. (This has the ARIMA(1,1,0) model as a special limiting case.)

- The smaller population 2 is modelled indirectly using the spreads in the period and cohort effects:
  - The spread between period effects, $S_2(t) = \kappa^{(1)}(t) - \kappa^{(2)}(t)$, is modelled as an AR(1) process with, potentially, a non-zero mean-reversion level. Random innovations in the AR(1) process are correlated with the $\kappa^{(1)}(t)$ innovations.
  - The spread between cohort effects, $S_3(c) = \gamma^{(1)}(c) - \gamma^{(2)}(c)$, is modelled as an AR(2) process with, potentially, a non-zero mean-reversion level. Random innovations in the AR(2) process are correlated with the $\gamma^{(1)}(c)$ innovations.
- Random innovations in the bivariate period-effect processes are assumed to be independent of random innovations in the bivariate cohort-effect processes.

The equations for this model are presented in Appendix B, and for a fuller discussion of the model, see Cairns et al. (2011b). A key element of the model fitting process in Cairns et al. (2011b) is the use of Bayesian methods.[22] The approach starts by combining the statistical likelihood functions for the death counts and the time series of underlying period and cohort effects: especially important where one or both of the populations are relatively small. Additionally, Bayesian methods produce a full posterior distribution both for process parameters (ν_1, ν_2, ψ, $C^{(2)}$, ζ_1, δ_1, ζ_2, ϕ_{11}, ϕ_{12}, ϕ_{21}, ϕ_{22}, $C^{(3)}$) and for historical values of the age, period and cohort effects.[23] The posterior distribution can then be used in a natural way to analyse the impact of parameter uncertainty on the results of our present analysis.

5 Simulation and valuation

5.1 Simulation

Simulation involves the following stages:

- First, in the case where we assume the parameters are unknown, we draw at random from the posterior distribution for the process parameters and for the historical age, period and cohort effects.

[22] For further discussion of mortality model fitting using Bayesian methods, see Czado et al. (2005), Pedroza (2006), Kogure et al. (2009), Reichmuth and Sarferaz (2008), and Kogure and Kurachi (2010).

[23] It has been demonstrated elsewhere (Cairns et al., 2006) that the inclusion of parameter uncertainty in process parameters can have a significant impact on forecast levels of uncertainty in future mortality rates.

- Next, we use simulation to extend the historical sequences of period and cohort effects by T years using the time series model discussed in Section 4. This then allows us to calculate the *underlying* death rates, $m_k(t,x)$, for years $t = 1,\ldots,T$ using equation (6).

- Finally, in experiments where we wish to take individual Poisson risk into account, we need to specify exposures and simulate death counts . Thus, we need to define what the exposures, $E_k(t,x)$, are for $t = 1,\ldots,T$, and then to simulate numbers of deaths using the conditional Poisson assumption:[24] that is,
$$D_k(t,x)|m_k(t,x) \sim \text{Poisson}\left(E_k(t,x).m_k(t,x)\right).$$

The output from the simulation step is, therefore, a set of deaths and exposures, rather than direct observation of the underlying death rates.

In the analysis that follows, we consider three cases that concern the specification of the exposures for the years 2006 to 2015:

- Case 1 (the large population or "no Poisson Risk" case). We set $E_k(t,x) = 100 \times E_k(0,x)$ for $k = 1,2$, $t = 1,\ldots,10$ and for all x.[25]

- Case 2 ("standard Poisson risk" case). We set the exposures for 2006 to 2015 to be equal to their 2005 levels: that is, $E_k(t,x) = E_k(0,x)$ for $k = 1,2$, $t = 1,\ldots,10$ and for all x.

- Case 3 ("small population Poisson risk" case). As case 2, but we set the exposures for population 2 for 2006 to 2015 to be equal to 0.001 times their 2005 levels: that is, $E_2(t,x) = 0.001 \times E_2(0,x)$, $t = 1,\ldots,10$ and for all x. This results in total exposures in each future year of just under 600, which might be considered typical for a medium sized pension plan considering hedging its longevity risk.

In all cases, exposures mostly decline with age from their 2005 values. However, we have not adjusted values to reflect cohorts of differing sizes, nor have we attempted to model reductions in the CMI exposures for reasons other than death, such as, policy maturities.

[24] For a discussion of the conditional Poisson assumption in a stochastic mortality context, see Brouhns et al. (2002) and Biffis et al. (2010). Li et al. (2009) put the case for a more-widely dispersed distribution than the Poisson. In a dynamic hedging context, the impact of Poisson risk has been considered previously by Dahl et al. (2008).

[25] The use of 100× is somewhat arbitrary, but is intended to be large enough that Poisson risk is very much less significant in the measurement of crude death rates. This makes the future CMI population much larger than the EW population, but even the latter has a small degree of Poisson risk. An alternative to the present version of case 1, that we have not tried, would be to set the observed number of deaths to be equal to its expected number, while leaving the exposures equal to those in case 2.

In case 1, the large population size should ensure that the observed death rates, $D_k(t,x)/E_k(t,x)$, are very close to the underlying death rates, $m_k(t,x)$, for $t = 1,\ldots,10$, and this should allow us to identify with precision the values of the underlying period and cohort effects in both a full or partial recalibration of the model. Case 2, in contrast, introduces greater noise in the death counts, resulting in less precision in those period and cohort effects that are estimated in 2015.

On average, the CMI male population has exposures that are about 10% of the size of the EW exposures. It follows that, at least under case 2, the Poisson risk will have a more noticeable impact on the CMI results.

5.2 Valuation

A theoretical value for $a_k(T,x)$ (compare with equation (1)) might be

$$a_k(T,x) = \sum_{s=1}^{\infty} P(T,T+s)p_k^{fwd}(T,s,x)$$

where

$$p_k^{fwd}(T,s,x) = E_Q\left[\left.\frac{S_k(T+s,x-T)}{S_k(T,x-T)}\right|\mathcal{M}_T\right],$$

$P(T,T+s)$ is the price at time T of the zero-coupon bond that pays 1 at time $T+s$ (which, here, we assume to be equal to $(1+r)^{-s}$) and $\mathcal{M}_t$ is the information provided about the development of mortality rates up to the end of year t. Expectations are taken with respect to a pricing measure Q. In a theoretical setting where the true model for mortality is known, when underlying death rates are observeable continuously, when there is a deep and liquid market, and where hedgers aim to minimise risk, Biffis et al. (2011) demonstrate how to identify Q uniquely from the wide range of possible choices. Here, we have chosen to equate Q to the real-world (or physical) probability measure, P. There are two reasons for this. First, our focus in this paper is on hedge effectiveness rather than on prices: prices are, of course, important, but they impact on different problems from the ones being considered here. Second, we believe that the shift from P to Q would not significantly change the correlation between the hedging instrument and the liability values.

For computational reasons, we will assume that the survival probabilities (now under measure P), $p_k^{fwd}(T,s,x) = E_P\left[S_k(T+s,x-T)/S_k(T,x-T)\middle|\mathcal{M}_T\right]$, can be approximated using a deterministic projection of mortality rates beyond time T rather than by taking the mean over the distribution of $S_k(T+s,x-T)$. The approximation used here is similar in spirit to those of Nielsen (2010), who examines Solvency II mortality stress tests, and Coughlan et al. (2011), who examine longevity hedging.[26] Note that the stochastic term $S_k(T+s,x-T)/S_k(T,x-T)$

[26] Alternative methods for approximating the expected survival probabilities have been proposed by Denuit et al. (2010), Cairns (2011a) and Dowd et al. (2010a, 2011b). Additionally, under certain conditions, one can explicitly model the two-dimensional forward survival probabilities to

equals $\exp\left[-\sum_{u=1}^{s} m_k(T+u, x+u-1)\right]$. We use, as a deterministic approximation to $m_k(T+s,x)$,

$$\hat{m}_k(T+s,x) = \exp\left[\beta^{(k)}(x) + n_a^{-1}\left(\kappa^{(k)}(T) + \nu_k s\right) + n_a^{-1}\gamma^{(k)}(T+s-x)\right] \qquad (7)$$

where $\beta^{(k)}(x)$, $\kappa^{(k)}(T)$ and $\gamma^{(k)}(T+s-x)$ are estimates of age, period and cohort effects that can be made using data up to time T, and ν_k is a population-k-specific drift in the period effect. In more general terms, valuation using deterministic projections is standard practice in the pensions industry, and it is this practice that we seek to emulate.

Assigning appropriate values to ν_1 and ν_2 in equation (7) is central to our analysis. We choose to equate ν_1 to the estimated drift in the random walk, $\kappa^{(1)}(t)$, made at time T, implying that $\hat{m}_1(T+s,x)$ is the median of the distribution of $m_1(T+s,x)$. The AR(1) model for the spread between $\kappa^{(1)}(t)$ and $\kappa^{(2)}(t)$ means that the median trajectory for $\kappa^{(2)}(t)$ is, in contrast, non-linear. However, in the long run, under the stochastic two-population model, the median trajectory of $\kappa^{(2)}(t)$ is asymptotically linear with gradient ν_1. Thus, with the linear approximation used in equation (7), an appropriate value to attach to ν_2 is also the drift of the random walk, $\kappa^{(1)}(t)$, to ensure consistency between forecasts of the two populations' mortality rates: that is, $\nu_1 = \nu_2$.[27]

Using equation (7) as an approximation, along with $\nu_1 = \nu_2$ as discussed above, we can approximate $p_k^{fwd}(T,s,x)$ by

$$\hat{p}_k^{fwd}(T,s,x) = \exp\left[-\sum_{u=1}^{s} \hat{m}_k(T+s, x+s-1)\right].$$

Finally, with a constant interest assumption, we have

$$a_k(T,x) \approx \hat{a}_k(T,x) = \sum_{s=1}^{\infty}(1+r)^{-s}\hat{p}_k^{fwd}(T,s,x). \qquad (8)$$

5.3 Calibration of the valuation model

Evaluation of equation (8) requires knowledge of: $\beta^{(k)}(x), \beta^{(k)}(x+1), \ldots$; $\kappa^{(k)}(T)$; ν_1; and the single cohort effect, $\gamma^{(k)}(T-x+1)$. The questions, therefore, arise as to how and when we estimate these various inputs. We consider three cases: full parameter certainty; partial parameter certainty; full parameter uncertainty.

---

avoid the need for approximations (Bauer et al., 2010). The method used here delivers accurate results using a simpler-to-implement algorithm.

[27] A more sophisticated approach would allow for the initial drift of $\kappa^{(2)}(t)$ to differ from ν_1, but then revert to ν_1 in the long run. Thus, in the expression for $\hat{m}_k(T+s,x)$, we might replace $\nu_2 s$ by $\nu_1 s + (\nu_2 - \nu_1)(1-\phi^s)/(1-\phi)$ where $\phi > 0$ is the AR(1) parameter in the spread between $\kappa^{(1)}(t)$ and $\kappa^{(2)}(t)$.

In all three cases, we will calibrate the valuation model using the single-population version of the APC model given in equation (6) above. This model is fitted separately to each of the EW and CMI datasets, and includes simple time series assumptions about the time series properties of the period and cohort effects. Further details on this are given in Appendix G.

In the full parameters-certain (PC) case, we proceed in the following steps:

- PC1: Fit the one-population model to each of the EW and CMI datasets running from 1981 to 2005: this is referred to as the *initial calibration* and is required for PC4.
- PC2: Fit a random walk model to the fitted period effect (EW) for 1981 to 2005. This gives us the estimated random-walk drift, ν_1^I.
- PC3: Simulations from 2005 to 2015 are carried out using the PC version of the two-population model (see Cairns et al., 2011b, for details).
- PC4: For each stochastic scenario taking us from 2005 to 2015: refit the one-population model to each population, subject to the constraint that age, period and cohort effects already estimated in the initial calibration remain unchanged. This means that we estimate only the 10 most recent period and cohort effects.
- PC5: For each scenario, annuity valuation at $T = 10$ requires projection of the period effects only, and so we use $\kappa^{(k)}(T)$ resulting from the time-T calibration, and the random-walk drift, ν_1^I, that was already estimated at time 0.

In the partial-parameters-certain (PPC) case, we proceed as follows:

- PPC1 to PPC4: Same as PC1 to PC4.
- PPC4A: Recalibrate the random-walk parameter values for the single-population period effect, $\kappa^{(k)}(t)$, using the W most recent values: in particular, we recalibrate the drift parameter,

$$\nu_1 = \left(\kappa^{(1)}(T) - \kappa^{(1)}(T - W)\right)/W.$$

  This is in contrast with the PC case, where ν_1 is left equal to its initial calibration, ν_1^I.
- PPC5: For each scenario, annuity valuation at $T = 10$ requires projection of the period effects only, and so we use $\kappa^{(k)}(T)$ resulting from the time-T calibration, and the recalibrated random-walk drift, ν_1.

In the full parameters-uncertain (PU) case, we proceed as follows:

- PU1/2: Not required.

Annuity price input variable	Case		
	PC	PPC	PU
$\kappa^{(k)}(T)$	Y	Y	Y
ν_1	N	Y	Y
$\gamma^{(k)}(T-x+1)$ higher ages, x	N	N	y
$\gamma^{(k)}(T-x+1)$ lower ages, x	Y	Y	Y
$\beta^{(k)}(y),\ y=x,x+1,\ldots$	N	N	y

Table 4: Input factors as a source of risk in the calculation of the annuity price, $a_k(T,x)$. N: no, the variable is fixed at time $t=0$ (end 2005). Y: yes, variable is not known until time T, and is a significant source of risk. y: the variable can be estimated at $t=0$, but is also subject to estimation uncertainty, and is subject to modest amounts of re-calibration risk at T.

- PU3: Simulations from 2005 to 2015 are carried out using the the PU version of the two-population model (see Cairns et al., 2011b, for details).
- PU4: For each stochastic scenario taking us from 2005 to 2015, use the historical-plus-simulated deaths and exposures to carry out a full recalibration (in contrast with partial recalibration in PC4 and PPC4) of the single-population APC models to the EW and CMI populations using actual deaths and exposures over a window of $W+1$ years (i.e., calendar years $T-W$ to T).
- PU4A: Recalibrate the random-walk parameter values for the single-population period effect, $\kappa^{(k)}(t)$, using the W most recent values: in particular, we recalibrate the drift parameter,
$$\nu_1 = \left(\kappa^{(1)}(T) - \kappa^{(1)}(T-W)\right)/W.$$
- PU5: For each scenario, annuity valuation at $T=10$ requires projection of the period effects only, and so we use $\kappa^{(k)}(T)$ resulting from the time-T calibration, and the recalibrated random-walk drift, ν_1.

5.4 The recalibration window, $W+1$

In the PPC and PU cases, ν_1 uses a recalibration window of W years up to time T to estimate ν_1. In this paper, we will assume in most of our numerical experiments that $W+1=20$ years. However, we will later discuss the sensitivity of the results to the choice of W.

5.5 Sources of uncertainty in $a_k(T,x)$

At the beginning of this section, we identified the various inputs required for the calculation of $a_k(T,x)$. We now consider which of these inputs causes uncertainty in $a_k(T,x)$ (see, also, Table 4):

- $\kappa^{(k)}(T)$ constitutes the principal source of randomness in $a_k(T,x)$. It is the only source of uncertainty in the full PC case for all but the lowest ages at time T (i.e., ages at time T with cohort effects that had not been estimated at time 0).

- In some cases, the value of $\gamma^{(k)}(T+1-x)$ used in the calculation of $a_k(T,x)$ is uncertain. Specifically, this is the case for younger ages, x, starting with the cohort aged x_0+T at the end of year T (where $x_0=50$ is the youngest age in our data) down to the cohort aged x_0 at the end of year T. None of these cohorts was included in the dataset available at time 0 (i.e., 2005), and so the relevant value of $\gamma^{(k)}(T+1-x)$ is uncertain and not measurable until some years later. See Appendix D, for further discussion.

- In the full PC case, there are no further sources of uncertainty. At time T, only the T most recent period and cohort effects are estimated; parameters already estimated at time 0 are left as they are; and the value of ν_1 is left unchanged from its initial calibration at time 0.

- In the PPC case, the estimated age, period and cohort effects are treated as known and not subject to parameter estimation uncertainty. In contrast with the full PC case, however, the random walk drift, ν_1, is recalibrated at time T, based on estimates of the period effect $\kappa^{(1)}(t)$ up to time T, and this means that ν_1 in the calculation of the $a_k(T,x)$ is uncertain.

- In the PU case, individual sample paths take account of parameter uncertainty in the 2005 calibration and the model is fully recalibrated at time T. Thus, besides the process risk inherent in the period up to time T and cohort effects for younger ages, full recalibration at T plus PU at $t=0$ means that the $\beta^{(k)}(x)$ and $\gamma^{(k)}(c)$ inputs to $a_k(T,x)$ are also uncertain.

6 Decomposing hedge effectiveness in customised and index longevity hedges

Basis risk and, therefore, hedge effectiveness are influenced by the risk factors outlined in section 1.1 above. We will now examine what we believe to be the most important risk factors that impact on the hedge effectiveness of longevity hedges, namely population basis risk, cohort effect uncertainty, recalibration risk, recalibration window, parameter uncertainty, and Poisson risk. We do this using the example of a pension plan that is considering employing either a customised or index value hedge as part of an asset-liability management exercise.

6.1 Correlation results for individual risk factors

We now take a detailed look at how the correlation between the liability and the hedging instrument values changes in response to the inclusion or exclusion of the various factors listed in the previous section. To recap: our liability value is $L(T) = a_2(T, x)$, where $T = 10$ (2015) and $x = 65$, and our hedging instrument value is $H(T) = a_k(T, y) - \hat{a}_k^{fxd}(0, T, y)$ where, again, $T = 10$, but y can range from 50 to 89, and the reference population might be either $k = 1$ (index-based hedge) or $k = 2$ (customised hedge).

In Figures 1 to 7, we investigate the impact on the correlation between liability and hedge values of: population basis risk (i.e., using an index hedge rather than a customised hedge); the inclusion of cohort effect uncertainty; the inclusion of recalibration risk in 2015; the length of the calibration window; the inclusion of parameter uncertainty in the 2005 calibration; and the inclusion of Poisson risk. In all of the figures, we plot $Cor(L(T), H(T))$ as a function of the hedging instrument reference age, y. In total, Figures 1 to 7 cover 14 experiments (A to N) that are outlined in Table 5.[28] [29] We are primarily interested in the effectiveness of index hedges, although, in most cases, we also plot the equivalent correlation curve for a customised hedge allowing us to compare the impact of the various risk factors on each type of hedge.

To help with the interpretation of the results, it is useful to consider a linear approximation of the annuity price. First, note that $a_k(T, x) = f(\beta^{(k)}[x], \kappa^{(k)}(T), \gamma^{(k)}(T - x + 1), \nu_1)$, where $\beta^{(k)}[x]$ is the column vector of age effects from age x upwards, $(\beta^{(k)}(x), \ldots, \beta^{(k)}(\omega))'$, ω is the maximum age, and $f(\cdot)$ is the annuity function governed by the deterministic projection of the period effects. The linearisation is then simply:

$$\begin{aligned} a_k(T, y) \quad \approx \quad & f(\hat{\beta}^{(k)}[x], \hat{\kappa}^{(k)}(T), \hat{\gamma}^{(k)}(T - x + 1), \hat{\nu}_1) \\ & + b_1'(y)\Big(\beta^{(k)}[x] - \hat{\beta}^{(k)}[x]\Big) + b_2(y)\Big(\kappa^{(k)}(T) - \hat{\kappa}^{(k)}(T)\Big) \\ & + b_3(y)\Big(\gamma^{(k)}(T - x + 1) - \hat{\gamma}^{(k)}(T - x + 1)\Big) + b_4(y)\Big(\nu_1 - \hat{\nu}_1\Big). \quad (9) \end{aligned}$$

This linearisation turns out to be a very accurate approximation to $f(\cdot)$, even with full PU and uncertainty in all of the $\beta^{(k)}[x], \kappa^{(k)}(T), \gamma^{(k)}(T - x + 1)$, and ν_1.

Turning now to the experiments listed in Table 5:

---

[28] Experiments F and H are listed here for completeness, but are not discussed below as they do not reveal anything in addition to the points already being made.

[29] The correlations plotted in these Figures are all estimates based on 1000 independent simulations of $L(T)$ and $H(T)$. As discussed in Appendix F, the correlations that are plotted have sufficiently small standard errors that none of the conclusions that we draw in the following sections are likely to be influenced by simulation errors. Wherever possible, we use *coupled* simulations: for example, all PC experiments will use the same 1000 sample paths for the underlying mortality rates up to time T. This means, for example, that the 1000 simulated values of $L(T)$ in one experiment are highly correlated with the 1000 simulated values of $L(T)$ in another experiment. This ensures greater robustness when we compare the results of different experiments.

Experiment	$H(T)$ Reference Population	Population Basis Risk	Parameter Estimation	Recalibration Risk	Recalibration Window	Poisson Risk
A	CMI	N	PC	N	20	N
B	CMI	N	PPC	Y	20	N
C	CMI	N	PU	Y	20	N
D	CMI	N	PU	Y	20	Y
E	EW	Y	PC	N	20	N
F	EW	Y	PC	N	20	Y
G	EW	Y	PPC	Y	20	N
H	EW	Y	PPC	Y	20	Y
I	EW	Y	PU	Y	20	N
J	EW	Y	PU	Y	20	Y
K	EW	Y	PU	Y	20	Y
L	EW	Y	PU	Y	35	Y
M	EW	Y	PPC	Y	35	N
N	EW	Y	PU	Y	20	Y*

Table 5: Key risk factors influencing the correlation between liability and hedge values for experiments A to N. The cohort effect, as a source of risk at younger ages, is present in all experiments. All experiments involve life annuities apart from experiment K which uses a temporary annuity that ceases at age 90. Experiments D, F, H, J, K and L involving Poisson risk relate to case 2 (Setion 5.1). (Y*) In Experiment N involving Poisson risk, the population 2 exposures (case 3) are 0.1% of the 2005 CMI exposures.

- **Benchmark customised hedge:** To provide a reference point, we start with a benchmark customised hedge (Figure 1). We take the simplest case, namely full parameter certainty (PC) without Poisson risk. The correlation curve (A) has two distinct parts to it. At ages 61 and above, the correlation is both very flat and very close to 1. In the PC case, $L(T)$ and $H(T)$ have $\kappa^{(2)}(T)$ as their single source of randomness, so the correlations are almost 1 (”almost” because there are still some slight non-linearities).[30]

- **Cohort effect uncertainty:**[31] Also in Figure 1, we note that the correlations drop away below age 61. This is because $a_2(T, y)$ also depends on the cohort effect $\gamma^{(2)}(c)$ for year of birth, $c = T - y + 1$. If the hedging instrument reference age, y, is less than 61, then the relevant value of $\gamma^{(2)}(c)$ is not known until after 2005 and therefore provides an additional source of randomness in $H(T)$. As we move from age 61 to younger ages (i.e., later years of birth), uncertainty in $\gamma^{(2)}(c)$ grows and, therefore, makes an increasing contribution to the overall risk in $H(T)$. Since this additional risk is not correlated with

[30]In this experiment, only $\kappa^{(2)}(T)$ is uncertain in the linearised equation (9), so the correlations between $a_2(T, x)$ and $a_2(T, y)$ in the linearised version for $x \neq y$ must be exactly equal to 1.

[31]Note that the caption to Figure 1 refers to *knowable cohort effects*. These refer to cohorts for which we do have data, but we choose to ignore these data because we have too few observations to be able to make reliable estimates of the cohort effect. See Appendix E and Cairns et al. (2009) for further discussion.

$\kappa^{(2)}(T)$, the correlation between $H(T)$ and $L(T)$ falls in line with the proportional contribution of $\gamma^{(2)}(c)$ to the uncertainty in $H(T)$.

- **Population basis risk:** In Figure 2, we introduce population basis risk by switching to the use of hedging instruments linked to the EW males population. We see that the broad impact of this switch is to pull down the correlation curve at all ages. Experiment E (dot-dashed line) gives correlations in the full PC case. As with curve A, curve E is fairly flat above age 61, reflecting the near-linear dependence of $L(T)$ and $H(T)$ on their single sources of risk, $\kappa^{(2)}(T)$ and $\kappa^{(1)}(T)$, respectively. This dependence is confirmed by the fact that $Cor(L(T), H(T)) \approx Cor(\kappa^{(1)}(T), \kappa^{(2)}(T))$ above age 61.

- **Recalibration risk:** Figure 3 shows the impact of model recalibration risk in the PPC case for both the customised (A to B) and index (E to G) hedges. First, consider customised hedges. This introduces a fresh source of risk, ν_1, into the calculation of annuity values. In experiment B (solid curve), above age 61, there are two distinct sources of risk ($\kappa^{(2)}(T)$ and ν_1, which, as previously discussed, is a linear function of $\kappa^{(1)}(T)$). Over the 61+ age range, correlations are still high, but, as y increases above age 65, correlations drift down gradually (curve B). For y close to age 65, $L(T)$ and $H(T)$ are exposed to the $\kappa^{(2)}(T)$ and ν_1 risks in approximately the same proportions (i.e., the ratio of $b_2(y)$ to $b_4(y)$ in equation (9)). However, as the reference age, y, increases, the relative impact of $\kappa^{(2)}(T)$ and ν_1 on $a_2(T, y)$ changes (i.e., $b_2(y)$ to $b_4(y)$), causing correlations to drop a little (solid line (B), right-hand end).

  Below age 61, in experiment B, there are three sources of risk: $\kappa^{(2)}(T)$, ν_1 and $\gamma^{(2)}(T - y + 1)$. The curve drops away as we reduce y for similar reasons as in experiment A. However, it is obvious that correlations for these lower ages are much higher in experiment B compared with A. In experiment B, $L(T)$ and $H(T)$ have, in absolute terms, significantly more risk than A, through additional uncertainty in ν_1. However, in relative terms, $L(T)$ and $H(T)$ have a much stronger dependence on common sources of risk ($\kappa^{(2)}(T)$ and ν_1) in experiment B than in experiment A and this results in a much higher correlation.

  Now consider the impact of recalibration risk on an index hedge. As a source of risk, ν_1 is common to both $L(T)$ and $H(T)$ over all reference ages. The inclusion of recalibration risk significantly increases the uncertainty in $L(T)$ and $H(T)$, but this is a perfectly correlated additional risk.[32] Thus, the impact of including PPC model recalibration risk is to increase the correlations and so raise curve E (dot-dashed line) significantly to curve G (solid line).

  Finally, in Figure 3, we compare experiments E (PC) and G (PPC) below age 61. In the PPC case (G), at lower ages, the additional risk in the cohort

[32] Referring to equation (9), $L(T)$ is approximately a linear combination of $\kappa^{(2)}(T)$ and ν_1, while $H(T)$ is a linear combination of $\kappa^{(1)}(T)$ and ν_1. In the PPC case, ν_1 is a risk that is common to both $L(T)$ and $H(T)$ and so raises the correlation between the two relative to the PC case, where ν_1 is fixed.

effect (Figure 8) is just as large in absolute terms as the full PC case (E), but, in relative terms, it contributes much less, because of the inclusion of the additional risk linked to ν_1 that is common to both $L(T)$ and $H(T)$. As a result, the decline in correlations below age 61 is less in the PPC case (G).

- **Recalibration window:** In Figure 4, we focus on the sensitivity of results to the choice of recalibration window. In experiment G, we use a 20-year window and, in experiment M, we use a 35-year window. Recall that $\nu_1 = (\kappa^{(1)}(T) - \kappa^{(1)}(T-W))/W$, so $W+1 = 35$ rather than $W+1 = 20$ reduces uncertainty in ν_1. Comparing experiments G and M, both $L(T)$ and $H(T)$ are less risky under M, because ν_1 is less risky. However, the correlation between $L(T)$ and $H(T)$ is also now lower because of their greater dependence, in relative terms, on the imperfectly correlated $\kappa^{(2)}(T)$ and $\kappa^{(1)}(T)$.

- **Parameter uncertainty:** Figure 5 adds in the impact of parameter uncertainty (PU) (experiments C and I, dashed lines). Introducing PU creates additional uncertainty in the process parameters (e.g., ν_1) and also in the latent state variables (the age, period and cohort effects). This additional uncertainty can be thought of as noise on top of the main signal and the noise added to the different components of $L(T)$ and $H(T)$ will be largely uncorrelated.[33] This creates additional risk that is mostly non-hedgeable (with the exception of age 65, where $L(T)$ and $H(T)$ refer to the same cohort) and so leads to lower correlations and lower hedge effectiveness.

  Figure 5 also shows the impact of moving from the PPC (curve G) to the PU case (curve I, dashed line) on the effectiveness of an index hedge. The impact is relatively small, with a magnitude that is similar *at most ages* to the customised hedge (experiments B and C). However, in contrast with the shift from B to C, we find here that there is no advantage to using a hedging instrument that is linked to exactly the same birth cohort as the liability being hedged: this reflects a lack of correlation in the PU setting between our estimates of the cohort effects in the two populations, $\gamma^{(2)}(T-x+1)$ and $\gamma^{(1)}(T-x+1)$ for $x = 65$.

- **Poisson risk:** In Figure 6, we look at the impact of adding in Poisson risk (experiments D and J, dotted lines, and N, dot-dashed line). In general, this should add to the uncertainty in both $L(T)$ and $H(T)$.[34] The impact on correlation of Poisson risk (C to D) is broadly similar in magnitude to the impact of PU (B to C). However, the impact does seem to vary with age. The reasons for these variations are not clear, suggesting that the impact of Poisson risk on the various inputs to $a_2(T,y)$ is complex. In contrast (experiments I

[33]For example, the noise added to $\kappa^{(1)}(T)$ and $\kappa^{(2)}(T)$ will have a low correlation, and, for $c_0 \neq c_1$, the noise added to $\gamma^{(2)}(c_0)$ and $\gamma^{(2)}(c_1)$ will also have a low correlation.

[34]Specifically, Poisson risk will perturb the input values in $f(\beta^{(k)}[x], \kappa^{(k)}(T), \gamma^{(k)}(T-x+1), \nu_1)$ in equation (9), and the fact that the noise is conditionally independent from (t,x) cell to cell means that these perturbations are likely to increase total risk.

and J), if we use an index-based hedge, then the gap between the Poisson and no-Poisson correlation plots seems to be reasonably uniform across all ages.

As might be expected, the impact on correlations for index hedges when we introduce Poisson risk depends significantly on the size of the underlying population. This is illustrated in Figure 6 as we move from experiment I (large population) through J (CMI-sized) to N (0.1% of CMI). The difference between I and J is modest. The additional Poisson risk in N significantly increases uncertainty in our estimate of the 2015 base table[35] that we calibrate for population 2 and this lowers the correlation between $L(T)$ and $H(T)$. But $L(T)$ and $H(T)$ are still just as reliant on the the common drift parameter ν_1 and so the impact of the additional Poisson risk is much less than it would be if the two populations were totally independent.

We considered four variants on the size population 2 to assess at what level Poisson risk starts to have a significant impact: 100% of the CMI (experiment J), 10%, 1% and 0.1% (experiment J). With a hedge instrument linked to age 65 in popluation 1, these resulted in correlations of 0.937, 0.923, 0.865 and 0.665 respectively. So we can see that there is a gradual shift in correlation as the population size falls. However, these figures suggest that it is only when the population size falls below, say, 10,000 members above age 50 that the measure of correlation begins to fall more rapidly: that is, between 1% and 2% of the CMI.

- **Term of the annuity:** In Figure 7, we make two further comparisons. As our baseline, we conduct experiment J which includes all risk factors. We investigate how sensitive the correlations are to the term of the annuity that underpins our calculations. Experiment J involves the use of a life annuity. In contrast, in experiment K, both $L(T)$ and $H(T)$ involve temporary annuities that cease at age 90. We can see that this lowers the correlations relative to curve I. This is explained by the fact that the recalibrated ν_1 has relatively little influence over short-dated cashflows and much greater influence over long-dated cashflows. By shifting to temporary annuities, we have therefore reduced the influence of ν_1 on $L(T)$ and $H(T)$, with a resulting lowering in the correlations.

  A similar shift would occur if we switched from a life annuity starting at age 65 to, say, a life annuity starting at age 75, since, on average, the annuity will be payable for less time.

[35] In each of the 1000 simulations, the true base table for population 2 in Experiment N in 2015 is, in reality, just the same as the true base table in Experiment I with no Poisson risk. However, the significant Poisson risk in N means that there is significantly more uncertainty in our estimate of the base table. So the observed decrease in the correlation is actually due to increasing estimation error in 2015 rather than any other source of risk. Furthermore, the additional estimation error due to the inclusion of Poisson risk will be independent of the population 1 valuation model calibrations.

As outlined in Appendix G, the base table for 2015 for population 2 is calibrated using population 2 data only. When the population is relatively small it might, however, be advisable to fit population 2 jointly with population 1 using data up to 2015 (for example, as in Cairns et al.,2011a). This might reduce the estimation error in the 2015 base table reported above.

- **Recalibration window revisited:** In Figure 7, which shows the results of experiment L, we repeat the shift from a recalibration window of 20 to one of 35 years, this time in a full PU setting with Poisson risk (compare with Figure 4.) The difference between J and L can be seen to be similar to (but slightly larger than) the shift from G to M. Thus, the impact of a change in the recalibration window can be seen to be not especially sensitive to the inclusion or otherwise of the less important risk factors (full PU and Poisson risk).

6.2 Analysis

Our discussion above of the individual plots focused on the incremental impact of the different risk factors. Here we look at the bigger picture and assess the overall impact and significance of each, beginning with the most important.

- *Population basis risk* is clearly a very significant factor. However, its negative impact on correlation and hedge effectiveness is not, perhaps, as large as might seem at first glance.

- *Recalibration risk* is also a factor of potential significance, although its precise impact depends on the type of hedge. For index hedges, it results in substantially increased correlations and hence hedge effectiveness;[36] for customised hedges, it has a very modest, negative impact.

  Further, as demonstrated in Figure 7, the impact of recalibration risk will be greater[37] if both the liability and hedging instrument values (in an index hedge) depend more heavily on more distant longevity-linked cashflows. In relative terms, these cashflows are much more sensitive to changes in the random-walk drift ν_1.

- The impact of recalibration risk on correlations was found to be quite sensitive to the length of the *calibration window* (e.g., 20 years or 35 years): a longer calibration window lowers both correlation and hedge effectiveness.

- *Cohort effect uncertainty* can cause correlations to be pulled down if the liability and hedging instrument refer to different cohorts either by year of birth or by reference population. The impact is modest if both of the relevant cohort effects had been estimated at time 0 (2005), and is a result of uncertainty in the estimates of those state variables. The impact is much more significant if one or other of the relevant cohort effects could not be estimated in 2005, thereby introducing additional uncertainty in the calculation of the annuity price at T.

[36] Where allowance for recalibration risk does increase correlations and hedge effectiveness, this helps to explain why the negative impact of population basis risk is smaller than anticipated.

[37] That is, the increase in correlation when moving from ignoring recalibration risk to including recalibration risk (as in the change from curve E to curve G in Figure 3) will be bigger.

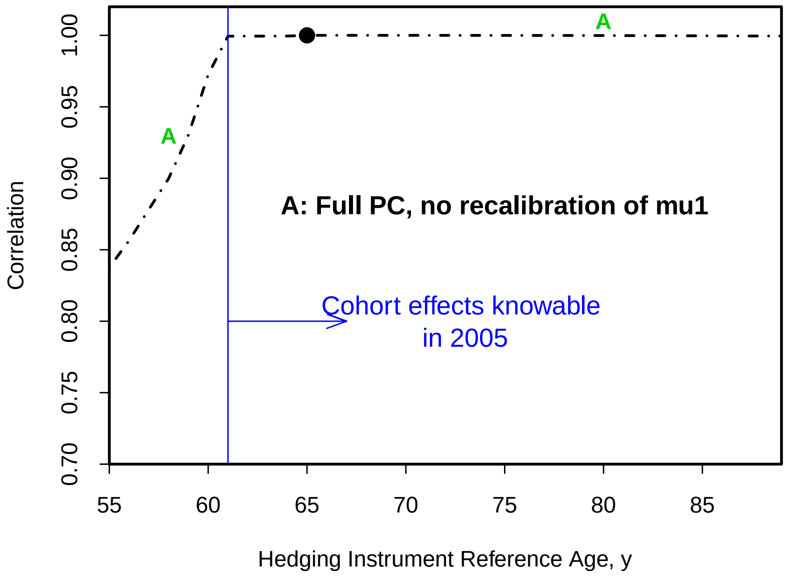

Figure 1: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiment A (Table 5) assumes a customised hedge (CMI reference population), full parameter certainty (PC) and no Poisson risk. Knowable cohort effects are discussed in Appendix D. The black dot identifies the liability reference age of 65.

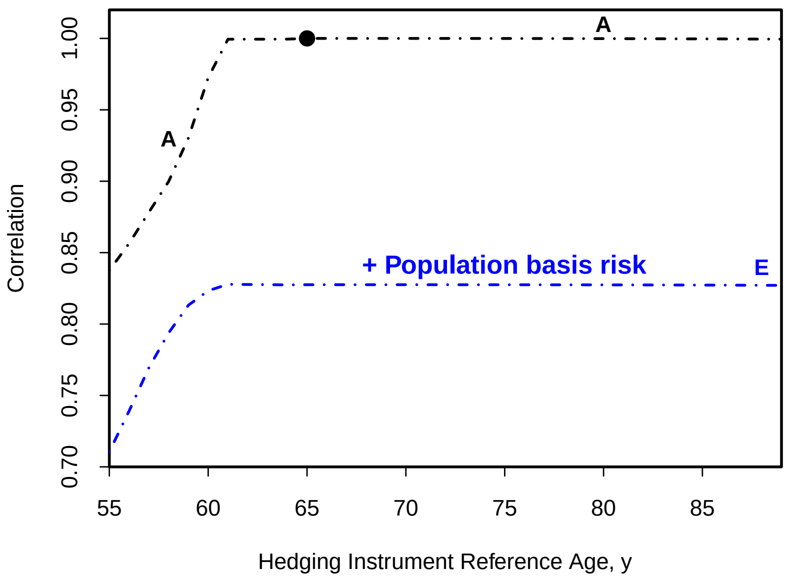

Figure 2: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments A, E (Table 5). A, E: full PC without Poisson risk. A: customised hedge. E: index hedge.

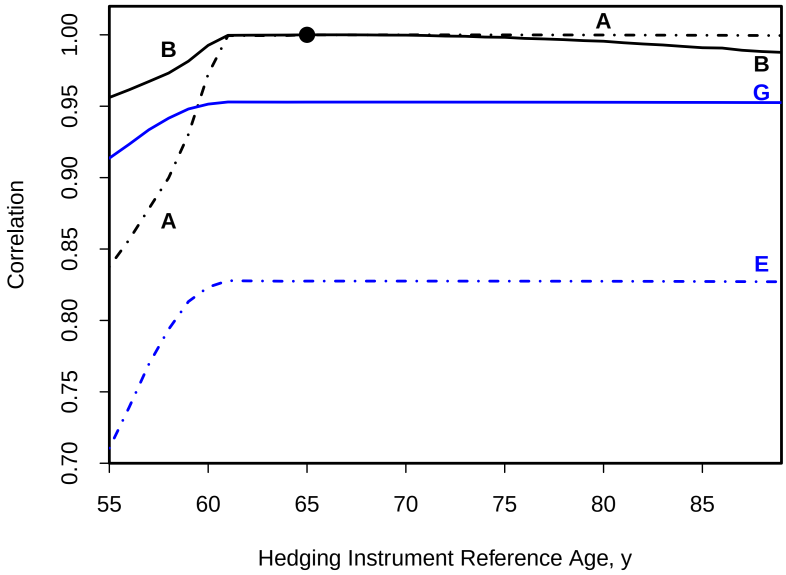

Figure 3: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments A, B, E, G (Table 5). A, B: customised hedges. E, G: index hedges. A, E: full PC. B, G: PPC with recalibration risk, 20-year recalibration window. All: without Poisson risk.

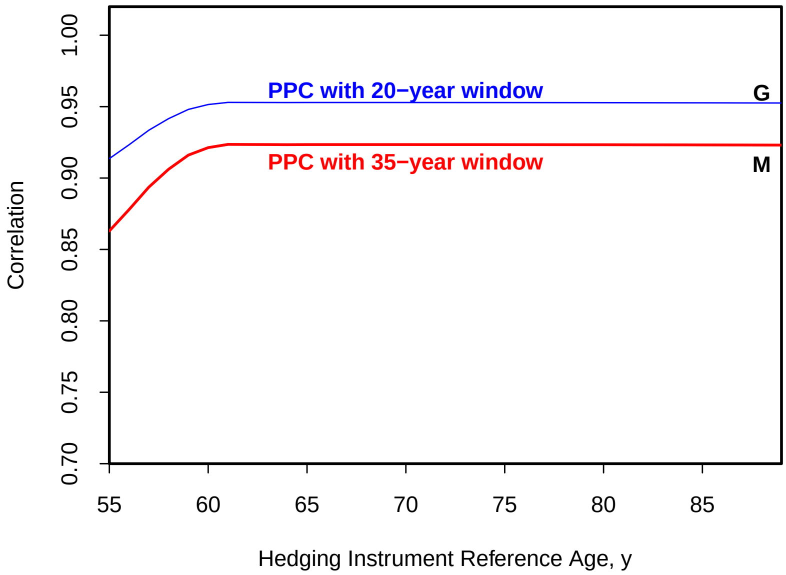

Figure 4: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments G, M (Table 5). G, M: index hedges, PPC. G: 20-year recalibration window. M: 35-year recalibration window.

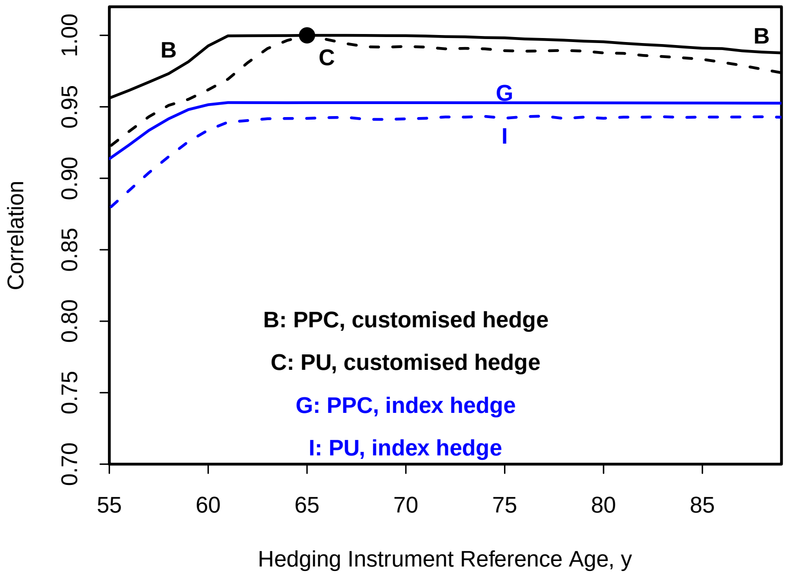

Figure 5: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments B, C, G, I (Table 5). All: without Poisson risk; 20-year recalibration window. B, C: customised hedges. G, I: index hedges. B, G: PPC. C, I: full PU.

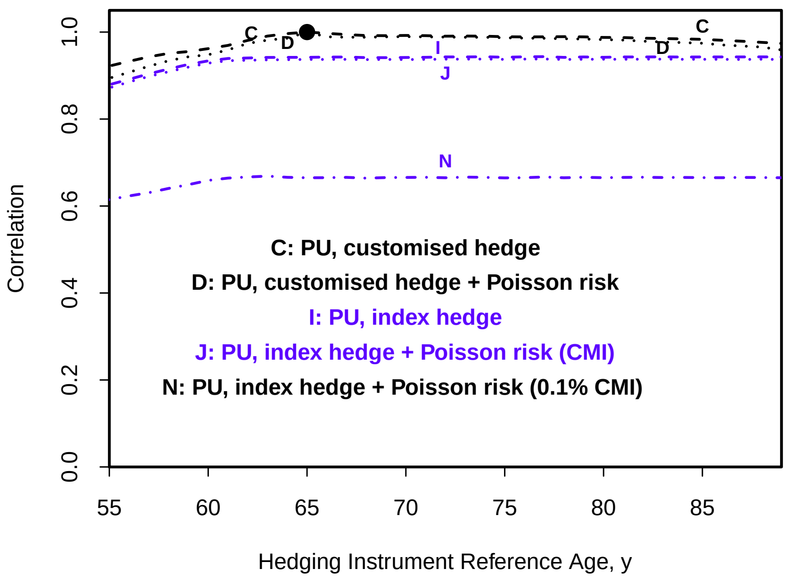

Figure 6: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments C, D, I, J, N (Table 5). All: full PU; 20-year recalibration window. C, D: customised hedges. I, J, N: index hedges. C, I: no Poisson risk. D, J: with Poisson risk (case 2). N with Poisson risk (case 3).

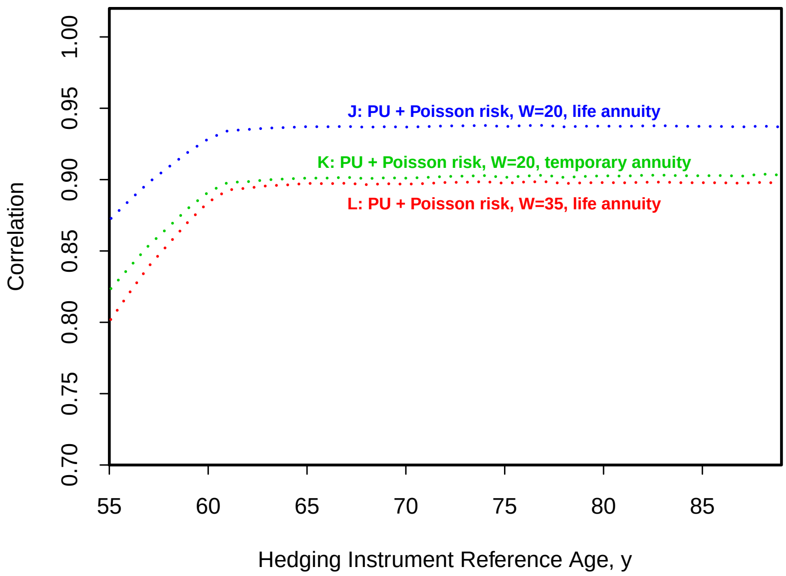

Figure 7: Correlation between the liability, $L(T)$, and hedging instrument, $H(T)$, values as a function of the hedging instrument reference age. Experiments J, K, L (Table 5). All: full PU with Poisson risk. J, K: 20-year recalibration window. L: 35-year recalibration window. J, L: life annuity. K: temporary annuity ceasing at age 90.

- Where we have already taken account of recalibration risk, the inclusion of other forms of parameter uncertainty (PU) and Poisson risk only have a modest impact on correlation and hedge effectiveness. However, where the hedger's own population is much smaller than the CMI population considered here (e.g., Experiment N), Poisson risk has a bigger impact.

7 Conclusions

This paper builds on the framework proposed by Coughlan et al. (2011) by analysing hedge effectiveness (with correlation as a proxy) using stochastic simulation (instead of historical bootstrapping). It is the first study to bring together, in a single stochastic modelling framework, the key risk factors influencing the effectiveness of longevity hedges, namely population basis risk, cohort effect uncertainty, recalibration risk, recalibration window, parameter uncertainty and Poisson risk.

To investigate longevity hedge effectiveness, we used a case study of a pension plan that wishes to hedge the value of its liability in 10 years' time to a male member who is currently aged 55. The liability is therefore equivalent to a deferred annuity. The plan had the choice of using either a customised hedge or an index hedge. We assumed, for the sake of illustration, that the mortality experience of the pension plan and the customised hedge was the same as the Continuous Mortality Investigation's male assured lives, while the mortality experience of the index hedge was the same as that for the England & Wales male population. For plans that aim to derisk, we argued that correlation between the value of plan liabilities and the value of a hedging instrument makes a good proxy as a measure of hedge effectiveness.

We found that population basis risk and uncertain future cohort effects are significant determinants of hedge effectiveness. However, we also showed that this was just the starting point. We discovered that correlation and hedge effectiveness are also affected to a significant extent by the inclusion of recalibration risk and the assumed length of the recalibration window. Beyond that, further sources of parameter uncertainty and Poisson risk (especially small populations) have a more modest, although still noticeable, impact. However, our analysis of the impact of Poisson risk suggest that we need to search for further ways to reduce estimation error in the calibration of small-population base tables, perhaps via joint modelling with much larger (e.g., national) populations.

The strong conclusion is that an analysis that ignores parameter uncertainty (including recalibration risk) might significantly underestimate the level of longevity risk, but, more importantly, might also underestimate the degree of hedge effectiveness of an index-based longevity hedge. So an unsophisticated and incomplete analysis of the problem might either lead to a decision not to hedge (because the level of risk is deemed not to be sufficiently high) or lead to a customised hedge being chosen in place of a cheaper index hedge (because the effectiveness of the latter has been underestimated). It is our belief that this specific conclusion would apply to a wider range of risk metrics than the simple use of variance (for example, expected shortfall or expected utility).

Our case study shows that longevity basis risk can be substantially hedged using index hedges as an alternative to customised longevity hedges. As a consequence, therefore, index longevity hedges – in conjunction with the other components of an ALM strategy – can provide an effective as well as a low cost alternative to a full buy-out of pension liabilities or even a strategy that involves the use of customised longevity hedges unless the population being hedged has fewer than around 10,000 members.

Apart from the hedging instrument reference population and reference age and the distinction between index and customised hedges, we have not investigated the impact on hedge effectiveness of the structure of the hedge. To do this, we would need to investigate such factors as the type of hedging instrument (namely, alternatives to a deferred longevity swap), the optimality and robustness of the hedge ratio (Cairns, 2011b), value versus cashflow hedges, static versus dynamic hedges, and the use of multiple hedging instruments, etc. Also omitted is an analysis of the impact of model risk: a substantial topic in its own right. Finally, we have not analysed the sensitivity of longevity hedge ratios to changes in the underlying assumptions. We leave these issues for future work.

References

Balistreri, E. J., and Hillberry, R. H. (2005) Estibration: An illustration of structural estimation as calibration. Working paper, Colorado School of Mines and University of Melbourne. (Downloaded 26/4/2012)

Bauer, D., Benth, F.E., and Kiesel, R. (2010) Modeling the forward surface of mortality. Working paper, University of Ulm. (Downloaded 30/5/2012)

Biffis, E., Denuit, M., and Devolder, P. (2010) Stochastic mortality under measure changes. *Scandinavian Actuarial Journal* 2010(4):284-311.

Blake, D., and Burrows, W. (2001) Survivor bonds: Helping to hedge mortality risk. *Journal of Risk and Insurance*, 68: 339-348.

Blake, D., Boardman, T., and Cairns, A.J.G. (2010) Sharing longevity risk: Why governments should issue longevity bonds. Pensions Institute working paper PI-1002 (accessed on 14/4/2011 at `http://www.pensions-institute.org/workingpapers/wp1002.pdf`).

Blake, D., Cairns, A.J.G., and Dowd, K. (2006) Living with mortality: Longevity bonds and other mortality-linked securities. *British Actuarial Journal*, 12: 153-197.

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

Cairns, A.J.G., Blake, D., and 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., and Dowd, K. (2008) Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal*, 2008(2-3): 79-113.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009) A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, 13: 1-35.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011a) 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., Coughlan, G.D., and Khalaf-Allah, M. (2011b) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41: 29-59.

Cairns, A.J.G. (2011a) Modelling and management of longevity risk: Approximations to survival functions and dynamic hedging. *Insurance: Mathematics and Economics*, 49: 438-453.

Cairns, A.J.G. (2011b) Robust hedging of longevity risk. Presented at Longevity 7: Seventh International Longevity Risk and Capital Markets Solutions Conference, Frankfurt, September 2011. *Working paper, Heriot-Watt University.*

Coughlan, G.D., Emery, S. and Kolb, J. (2004) HEAT (Hedge Effectiveness Analysis Toolkit): A consistent framework for assessing hedge effectiveness under IAS 39 and FAS 133. *Journal of Derivatives Accounting*, 1(2): 221-272.

Coughlan, G., Epstein, D., Sinha, A., and Honig, P. (2007) q-forwards: Derivatives for transferring longevity and mortality risk. Available at www.lifemetrics.com.

Coughlan, G.D. (2009). Longevity risk transfer: Indices and capital market solutions. In Barrieu, P.M. and Albertini, L. (eds), *The Handbook of Insurance Linked Securities*, Wiley, London.

Coughlan, G.D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A.J.G., Blake, D. and Dowd, K., (2011) Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, 15: 150-176.

Czado, C., Delwarde, A., and Denuit, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, 36: 260-284.

Dahl, M., Melchior, M., and Møller, T. (2008) On systematic mortality risk and risk minimisation with survivor swaps. *Scandinavian Actuarial Journal*, 2008(2-3): 114-146.

Dahl, M., Glar, S., and Møller, T. (2009) Mixed dynamic and static risk minimization with an application to survivor swaps. 19th International AFIR Colloquium, Munich, September 2009.

Denuit, M., Haberman, S., and Renshaw, A.E. (2010) Comonotonic approximations to quantiles of life annuity conditional expected present values: Extensions to general ARIMA models and comparison with the bootstrap. *ASTIN Bulletin*, 40: 331-349.

Detlefsen, K., and Härdle, W.K. (2007) Calibration risk for exotic options. *Journal of Derivatives*, 14: 47-63.

Dowd, K., Blake, D., and Cairns, A.J.G. (2010a) Facing up to uncertain life expectancy: The longevity fan charts. *Demography*, 47: 67-78.

Dowd, K., Blake, D., Cairns, A.J.G., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010b) Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, 47, 255-265.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., and Khalaf-Allah, M. (2011a) A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15: 334-356.

Dowd, K., Blake, D., and Cairns, A.J.G. (2011b) A computationally efficient algorithm for estimating the distribution of future annuity values under interest-rate and longevity risks. *North American Actuarial Journal*, 15: 237-247.

Jarner, S.F., and Kryger, E.M. (2011) Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, 41: 377-418.

Kogure, A., Kurachi, Y., and Kitsukawa, K. (2009) A Bayesian evaluation of longevity risk: Model comparison, measuring and pricing. Working paper, Keio University.

Kogure, A., and Kurachi, Y. (2010) A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, 46: 162-172.

Krugman, P. (2011) Calibration and all that (wonkish).
http://krugman.blogs.nytimes.com/2011/10/18/calibration-and-all-that-wonkish/
(Accessed 17/7/12.)

Lee, R.D., and Carter, L.R. (1992) Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87: 659-675.

Li, J.S.-H., and Hardy, M.R. (2011) Measuring basis risk involved in longevity hedges. *North American Actuarial Journal*, 15: 177-200.

Li, J.S.-H., Hardy, M.R., and Tan, K.S. (2009) Uncertainty in model forecasting: An extension to the classic Lee-Carter approach. *ASTIN Bulletin,* 39: 137-164.

Li, N., and Lee, R. (2005) Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, 42(3): 575-594.

Nielsen, L.H. (2010) Assessment of the VaR(99.5%) for longevity risk. Working paper, Sampension, Denmark.
`http://www.cea.eu/index.php?page=non-cea-publications`
(Accessed 3/8/2010)

Olivieri, A. and Pitacco, E. (2009) Stochastic mortality: The impact on target capital. *ASTIN Bulletin*, 39: 541-563.

Pedroza, C. (2006) A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics*, 7: 530-550.

Plat, R. (2009) Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, 45: 123-132.

Reichmuth, W. and Sarferaz, S. (2008) Bayesian demographic modelling and forecasting: An application to US mortality. SFB 649 Discussion paper 2008-052.

Wills, S. and Sherris, M. (2010) Securitization, structuring and pricing of longevity risk. *Insurance: Mathematics and Economics*, 46: 173-185.

Wylie, J.J., Zhang, , and Siu, T.K. (2010) Can expected shortfall and value-at-risk be used to statically hedge options? *Quantitative Finance*, 10: 575-583.

A Risk metric for hedge effectiveness

In Section 3.2, we introduce variance as our risk metric. This choice is consistent with the use of the term *hedge effectiveness* in international accounting standards (IAS39: see Coughlan et al., 2004). Although a metric is not actually specified in IAS39, two points that we might infer are as follows:

- We are concerned with *unanticipated* changes in the assets and liabilities and not the cost of setting up a hedge. This is consistent with the use of variance or standard deviation as risk metrics as obvious choices. It is also consistent with the use of value-at-risk or expected shortfall where these are defined relative to the median for the chosen hedging strategy (i.e., this ignores the cost of the strategy but is consistent with international accounting standards).
- With a value hedge (IAS 39 allows a focus on either value hedges or cashflow hedges) we are concerned with how much of a reduction is there in the volatility of the net financial position. This points to the use of a symmetrical risk metric such as variance.

Putting accounting standards to one side, there are other possibilities for assessing the effectiveness of a hedge including value-at-risk relative to the *unhedged* median, expected shortfall (see Wylie et al., 2010) and expected utility (see, for example, Appendix A of Cairns, 2011b, in a longevity risk setting). In these settings, the price of the hedge comes into play and so we are interested in value for money as well as risk reduction. We have chosen not to consider these possibilities here in order that we can focus attention on issues other than pricing, including the requirements of international accounting standards.

B Two-population mortality model: Stochastic model details

Define

$R_2(t) = \kappa^{(1)}(t),\ \ S_2(t) = \kappa^{(1)}(t) - \kappa^{(2)}(t),\ \ R_3(c) = \gamma^{(1)}(c),\ \ \text{and}\ \ S_3(c) = \gamma^{(1)}(c) - \gamma^{(2)}(c).$

Then

$$
\begin{aligned}
R_2(t+1) &= R_2(t) + \nu_1 + C_{211} Z_{21}(t+1) \\
S_2(t+1) &= \nu_2 + \psi(S_2(t) - \nu_2) + C_{221} Z_{21}(t+1) + C_{222} Z_{22}(t+1) \\
\tilde{R}_3(c) &= R_3(c) - \zeta_1 - \delta_1(c - \bar{c}) \\
\tilde{S}_3(c) &= S_3(c) - \zeta_2 \\
\tilde{R}_3(c+1) &= (\phi_{11} + \phi_{12}) \tilde{R}_3(c) - \phi_{11}\phi_{12} \tilde{R}_3(c-1) + C_{311} Z_{31}(c+1) \qquad (10) \\
\tilde{S}_3(c+1) &= (\phi_{21} + \phi_{22}) \tilde{S}_3(c) - \phi_{21}\phi_{22} \tilde{S}_3(c-1) + C_{321} Z_{31}(c+1) \\
&\qquad + C_{322} Z_{32}(c+1). \qquad (11)
\end{aligned}
$$

The details of these equations are as follows:

- $Z_{21}(t+1),\ Z_{22}(t+1),\ Z_{31}(c+1),\ Z_{32}(c+1)$ are i.i.d. standard normal random variables.
- ν_1 is the drift in the random walk $R_2(t)$.
- ν_2 and ψ are the mean-reversion level and the AR(1) parameter respectively of the period-effect spread, $S_2(t)$. For the process to be stationary (mean reverting), we require $-1 < \psi < 1$.
- Define $C^{(2)} = \begin{pmatrix} C_{211} & 0 \\ C_{221} & C_{222} \end{pmatrix}$, and $V^{(2)} = C^{(2)} C^{(2)\prime}$. $V^{(2)}$ is the 1-year-ahead conditional covariance matrix of $(R_2(t), S_2(t))'$.
- $\bar{c}$ is defined as $(c_0 + c_1 + 2)/2$, where (c_0, c_1) is the complete range of years of birth cohorts covered in the dataset.
- $\zeta_1 + \delta_1(c - \bar{c})$ is the linear trend, to which $R_3(c)$ is reverting.
- ζ_2 is the mean-reversion level of the cohort-effect spread, $S_3(c)$.
- $\tilde{R}_3(c)$ and $\tilde{S}_3(c)$ are AR(2) processes that are mean reverting to 0.
- Define $C^{(3)} = \begin{pmatrix} C_{311} & 0 \\ C_{321} & C_{322} \end{pmatrix}$ and $V^{(3)} = C^{(3)} C^{(3)\prime}$. $V^{(3)}$ is the 1-year-ahead conditional covariance matrix of $(R_3(t), S_3(t))'$ (and of $(\tilde{R}_3(t), \tilde{S}_3(t))'$).
- ϕ_{11}, ϕ_{12}, ϕ_{21} and ϕ_{22} are the AR(2) parameters for the processes $\tilde{R}_3(c)$ and $\tilde{S}_3(c)$. For the processes to be stationary, we require each of ϕ_{11}, ϕ_{12}, ϕ_{21} and ϕ_{22} to lie between -1 and $+1$.

C Why might the inclusion of parameter uncertainty strengthen correlations?

We present here a simple example. Suppose that L is our risky liability and H is the payoff on a hedging instrument.

- $L = \nu + \epsilon_L$, where the error ϵ_L has zero mean, and variance σ_L^2, and ν is the mean of L.
- $H = \nu + \epsilon_H$, where the error ϵ_H has zero mean, and variance σ_H^2, and ν is the mean of H.
- ϵ_L and ϵ_H are known to be independent.
- In the parameters-certain case, the correlation between L and H is zero.
- Now suppose that ν is subject to some estimation error, and that ν has mean $\hat{\nu}$ and variance σ_ν^2.
- When we include parameter uncertainty in our forecasts, we see that the correlation between L and H is now positive.

D Uncertainty in cohort effects

As remarked in the main text, for older cohorts, the value of the cohort effect, $\gamma^{(k)}(T - x - 1)$, to be used in the calculation of the annuity price $a_k(T, x)$, can already be estimated in 2005. This is illustrated in Figure 8.

The black dots and circles represent the data available in 2005. Individual cohorts follow diagonals moving in a north-easterly direction. This plot shows data for ages 50 to 89 from 1981 to 2005 and covers cohorts born in 1892 to 1955. In Figure 8, these are labelled on the right hand side as having a known or *knowable* cohort effect in 2005.

At the end of 2015, we seek to calculate annuity values for various cohorts. Cohorts born after 1955 fall in the lower right corner of Figure 8, and are characterised by the fact that the annuity value in 2015 will include a simulated value for the cohort effect, since this was not known in 2005.

In the software developed for a previous study (Cairns et al., 2009), we chose to exclude cohorts with four or fewer observations (or cells) to avoid overfitting. The four most recent cohorts (1952 to 1955), therefore, have knowable cohort effects in 2005, but we choose to ignore the very limited data that we have available in 2005 and leave estimation of their values until 2015. As a consequence, in the full parameters certain (PC) and partial parameters certain (PPC) cases, the cohort effects for the 1952 to 1955 cohorts (identified by the "knowABLE" diagonal band of

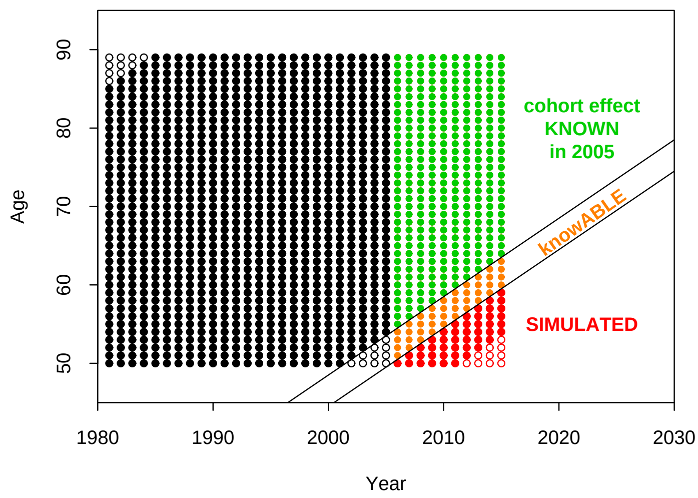

Figure 8: Timing of the estimation of the period and cohort effects in the parameters certain (PC) case. Black dots and circles: period effect estimated in 2005. Diagonal black and green dots: cohort effect estimated in 2005. Diagonal orange and red dots: cohort effect estimated only in 2015. Diagonal red circles: cohort effect not estimated even in 2015. Vertical green, orange and red dots and circles: period effect estimated in 2015. Black and red circles (lower right): cohort effects that are knowable in 2005 and 2015 respectively, but are not estimated due to limited quantity of data.

dots in Figure 8) are also not known until 2015 and therefore add to the uncertainty in the calculation of the annuity price. However, this uncertainty turns out to be negligible in the large-population version of the model.

E Different measures of correlation

Consider the following simple example of a two-period, two-factor model:

$$\begin{aligned} X_1 &= Z_1(1) && + Z_1(2) \\ X_2 &= \rho Z_1(1) + \sqrt{1-\rho^2} Z_2(1) && + Z_2(2) \end{aligned}$$

where the $Z_k(t)$ are independent and identically distributed standard normal random variables. $Z_1(1)$ and $Z_2(1)$ are known at time 1; $Z_1(2)$ and $Z_2(2)$ are known at time 2; and $\mathcal{F}_1$ represents the sigma-algebra generated by $(Z_1(1), Z_2(1))$, that is, the information available at time 1.

Let

$$\begin{aligned} L(1) &= E[X_2|\mathcal{F}_1] = Z_1(1) \\ H(1) &= E[X_1|\mathcal{F}_1] = \rho Z_1(1) + \sqrt{1-\rho^2} Z_2(1). \end{aligned}$$

The three measures of correlation that we might be interested in are

$$\begin{aligned} \mathrm{cor}(X_1, X_2) &= \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}} = \frac{1}{2}\rho && (12) \\ \mathrm{cor}(X_1, X_2|\mathcal{F}_\infty) &= \mathrm{cor}(Z_1(2), Z_2(2)) = 0 && (13) \\ \mathrm{cor}(L(1), H(1)) &= \mathrm{cor}\left(Z_1(1), \rho Z_1(1) + \sqrt{1-\rho^2} Z_2(1)\right) = \rho. && (14) \end{aligned}$$

The correlation defined in Equation (14) is what we are mainly concerned with in this paper for value hedging with a time horizon of 1 time unit.[38]

Equation (12) is relevant for cashflow hedging where we use X_1 as a hedging instrument to hedge the liability of X_2, with both payable at time 2.

F Estimates of the hedge effectiveness

Recall from equation (5) that the hedge effectiveness is given by

$$R^2(h) = \rho^2 \left(1 - \frac{(h-h^*)^2}{h^{*2}}\right) = \rho^2 \left(1 - \alpha^2\right)$$

where h^* is the theoretically optimal hedge ratio and $\alpha = (1 - h/h^*)$ provides a measure of the relative deviation between the actual hedge ratio and the optimum. Ideally, $h = h^*$, but the two might differ because of estimation error. Thus, suppose that $\alpha \sim N(\nu_\alpha, \sigma_\alpha^2)$, so that α^2 has a non-central chi-squared distribution with

[38] Additionally, the emphasis in the main paper over the choice of valuation model relates, here, to the choice of measure underpinning the conditional distribution at time 1 of $(Z_1(2), Z_2(2))$.

1 degree of freedom and non-centrality parameter ν_α^2. Standard properties of the non-central chi-squared distribution lead to the following results:

$$\begin{aligned} E[R^2(h)] &= \rho^2\left(1-\sigma_\alpha^2(1+\nu_\alpha^2)\right) \\ \text{and } Var[R^2(h)] &= 2\rho^4\sigma_\alpha^4(1+2\nu_\alpha^2). \end{aligned}$$

Provided $|\nu_\alpha| << 1$, it follows that (i) the difference between ρ^2 and $E[R^2(h)]$ is primarily determined by σ_α^2 (i.e., the additional impact of ν_α is negligible) and (ii) $Var[R^2(h)]$ is primarily determined by σ_α^4.

For example, for experiment I, the distribution of α was determined as follows in the case where we use the EW population with a reference age of 65 to hedge the age 65 CMI liability.

- Let $h^* = -\text{Cov}(H,L)/\text{Var}(H)$, where $\text{Cov}(H,L)$ and $\text{Var}(H)$ are the empirical covariance and variance based on the original sample of 1000 pairs of (H,L).
- Assume that this sample of 1000 pairs of (H,L) has been drawn from a bivariate normal distribution with mean vector, ν, and covariance matrix, V, determined by the simulated values of H and L: i.e., $\text{BVN}(\nu,V)$. The assumption of bivariate normality is consistent with the simulated data.
- Now carry out a parametric bootstrap as follows.

  For $i = 1,\ldots,100000$:

  - Simulate 1000 pairs of (H,L) from the $\text{BVN}(\nu,V)$ and calculate $h_i = -c_i/v_i$ where c_i is the empirical covariance between the simulated vectors H and L in this repetition, and v_i is the empirical variance of H.
  - Let $\alpha_i = 1 - h_i/h^*$.
- Analyse the empirical distribution of the α_i.

We found that:

- ν_α was indistinguishable from 0 with a standard error of 0.000036.
- $\sigma_\alpha^2 = 0.0001264$ with a standard error of 0.0000006.

From this experiment, we conclude that, since ν_α is very close to zero, and σ_α^2 is much smaller than ρ^2, the use of $R^2(h^*) = \rho^2$, where h^* is estimated from the simulated data provides an accurate approximation.

F.1 Standard error of the correlation estimate

From the same bootstrapping experiment, we can also calculate a standard error for the correlation ρ. Thus, the estimated value of ρ in experiment I at 0.942 is found to have a standard error of about 0.004.

G Single population APC model fitting

In our numerical experiments, our starting point was to use the APC model fitting procedure outlined in Cairns et al. (2009) with R code available online at `www.LifeMetrics.com`. This code maximises the Poisson log-likelihood function

$$l_1(\beta, \kappa, \gamma | D) = \sum_{t,x} D(t,x) \log m(t,x) - m(t,x)E(t,x) + \text{constant}$$

where $m(t,x) = \exp[\beta(x) + n_a^{-1}\kappa(t) + n_a^{-1}\gamma(t-x)]$ over the $\beta(x)$, $\kappa(t)$ and $\gamma(t-x)$.

However, we found that, for small populations, the possibility might arise that there are zero deaths in one or more calendar years across all ages or zero deaths amongst one or more cohorts during the periods of observation of these cohorts. In the former case, the log-likelihood is maximised by setting $\kappa(t) = -\infty$ for the relevant year, which is not realistic from a perspective of biological reasonableness.

It was seen as desirable that the fitted $\kappa(t)$ and $\gamma(t-x)$ for a small sub-population should exhibit similar time series characteristics (in particular, their levels of volatility from one year to the next) to the much larger EW population. Thus we chose to maximise the following Bayesian posterior distribution:

$$l(\beta, \kappa, \gamma, \theta | D) = l_1(\beta, \kappa, \gamma | D) + l_2(\kappa, \theta) + l_3(\gamma, \theta) + l_4(\theta),$$

where $l_2(\kappa, \theta)$ is the log-likelihood for the $\kappa(t)$ time series (assumed to be a random walk), $l_3(\gamma, \theta)$ is the log-likelihood for the $\gamma(t-x)$ time series (also, for simplicity, assumed to be a random walk), θ is the vector of parameters of the time series processes (i.e., the drifts and volatilities: $(\mu_\kappa, \sigma_\kappa, \mu_\gamma, \sigma_\gamma)$), and $l_4(\theta)$ is the Bayesian prior distribution for θ (a Normal-Inverse Gamma distribution for each pair of parameters).

In this extended model, the problem associated with zero deaths is now counterbalanced by the requirements in l_2, l_3 and l_4 for the $\kappa(t)$ and $\gamma(t-x)$ estimates to behave like random walks with the given parameters. Values of $-\infty$ would cause l_2 or l_3 to equal $-\infty$.

In our experiments, for CMI-sized and larger populations, there was no noticeable difference between the old fitting programme (i.e., maximising l_1 only) and the new (i.e., maximising the Bayesian posterior, l). Differences start to emerge when our population is about 10% or less of the CMI, and, in these circumstances, we find that estimates of $\kappa(t)$ and $\gamma(t-x)$ tend to be less volatile under the extended Bayesian paradigm.

As a word of caution, results reported for Experiment N in Figure 6 were found to be moderately sensitive to the median of the prior distribution for σ_κ (i.e., different medians that were considered to be realistic in the context of historical data up to 2005 would give correlations in a range of ± 0.04 around the value of 0.66 that we observe in Figure 6).