A methodology for assessing basis risk

Abstract of the London Discussion

[Institute and Faculty of Actuaries, Sessional Research Event, London, 8 December 2014]

This abstract relates to the following paper: IFoA/LLMA (2014) Longevity Basis Risk. *A methodology for assessing basis risk by Cass Business School and Hymans Robertson LLP.* Available at http://www.actuaries.org.uk/sites/all/files/IFoA%20LLMA%20Longevity%20Basis%20Risk%20Report_0.pdf

The Chairman (Mr R.P. Bugg, F.I.A.): This sessional paper event is about longevity basis risk. I work for Milliman and represent the Longevity Risk Basis Risk Working Group. The working group is a joint working group set up by the Institute and Faculty of Actuaries (IFoA) and the Life and Longevity Markets Association (LLMA). We were set up with the goal of producing tangible and useful output on longevity basis risk, which is accessible to all practitioners with an interest in the area.

We realised, however, that this is a far bigger and more complex question than could be answered by a working group with only limited time at its disposal. The group, therefore, secured funding and was pleased to engage Hymans Robertson and Cass Business School to carry out an initial phase of research. The purpose of this event is to allow Hymans Robertson and Cass to present the findings from this phase (IFoA/LLMA, 2014a, 2014b).

By way of background, the market for annuity and pension scheme de-risking has been growing year on year and 2014 has been the most active year to date. There are constraints on this market, for example, passing on the risk for deferred lives is still extremely difficult, and very few deals include deferred lives and the capacity and appetite of reinsurers for longevity risk is subject to some constraints. To tap into the full capacity of reinsurers and the capital markets, there may need to be a growth in so-called index-based hedges, where the payoff from the risk bearer relates to the mortality experience of a population that is different from the book in question. For example, one might use the UK population to determine the payoff. This then introduces basis risk into the hedge between the book population and the reference population. It is lack of understanding of basis risk that we see as one of the key barriers to the growth of a market in index-based transactions. Even outside the transactions area, the question of basis risk still arises. Longevity practitioners still almost universally set their trend assumptions by reference to the UK population. This introduces basis risk into their modelling against which they may wish to hold capital. Understanding and measuring basis risk is not an issue that is restricted to firms with pension schemes undertaking transactions.

I will now introduce our speakers for the evening. First is Steven Baxter from Hymans Robertson. Mr Baxter enjoys a diverse role, mixing strategic advice to insurers, pension schemes and banks, managing longevity risk and actively researching a wide range of longevity-related issues. His advisory roles include being lead longevity consultant on the second largest UK pension scheme longevity swap traded to date and providing longevity pricing advice to numerous large insurers

and reinsurers. He leads the longevity research and development programme at Hymans Robertson and is the architect of the longevity analytics within Club Vita, the longevity comparison club for occupational pension schemes. He has been leading the Hymans Robertson team on this project and is going to take us through the key conclusions of the project.

Following Mr Baxter will be Andrés Villegas and Pietro Millossovich from Cass Business School, who will take us through some of the detail and evidence behind the group's conclusions.

Mr Villegas is a PhD student in actuarial science, focussing on the modelling and projection of mortality. Before his doctoral studies, he obtained an MSc degree in industrial engineering from the Universidad de Los Andes in Colombia, and worked as a risk analyst at one of the biggest Colombian life insurance companies. His research interests include mortality modelling, longevity risk management and the application of optimisation techniques in actuarial science and finance.

Dr Millossovich is a senior lecturer in actuarial science at the Faculty of Actuarial Science and Insurance at Cass Business School. Previously, he was a lecturer at the University of Trieste in Italy. He holds a BSc in Statistics and Actuarial Science from the University of Trieste, a DEA in probability and finance from The University of Paris and a PhD in mathematics applied to decisions in economics and finance from the University of Trieste.

Then Mr Baxter will speak again, talking about the characterisation approach. Following that, we will have questions and answers. The person taking the lead in responding to questions and answers from the audience will be Andrew Gaches.

Mr Gaches is a longevity consultant with many years' experience, advising pension schemes, banks and insurers. His longevity expertise was central to the establishment of Club Vita. He now focusses on guiding clients through the process of recognising, quantifying and managing the longevity risks they face. He is a regular speaker at conferences, has written articles and authored papers on longevity, and is a long-standing member of the CMI Statistical Mortality Committee.

Mr S.D. Baxter, F.I.A. (introducing the paper): First I take this opportunity, on behalf of the whole team, to thank the IFoA and the LLMA for sponsoring this research. I will outline some of the key conclusions of our research. We have a case study which shows some of the results of applying our methodology in practice. We will also touch on some of the practicalities that you would face when using these methods, to help stimulate some discussion at the end of the presentation.

Seven key conclusions come out of our research. The first two of these support the previous work that the Longevity Basis Risk Working Group had done in identifying the key issues at hand. Firstly, there is a knowledge gap. We have an absence of knowledge about the quantum of basis risk. Secondly, an absence of knowledge brings with it a tools gap, that is, the absence of a well-defined modelling framework to quantify basis risk. Both of these observations were crucial to the sponsoring of this research.

Before I take you through our other five main conclusions, I want to give you the headline conclusion of what our modelling framework looks like. The framework consists of four key questions that users answer to identify an appropriate way to assess longevity basis risk as shown in the decision

462

tree on page 6 of the report (also figure 7.1) and page 6/7 of user guide. We believe that the majority of people using this decision tree will either:

- follow the route that goes across the top of the decision tree, a very specific model suitable for their circumstances, which is called the M7-M5 model; or
- use an alternative method, which we have called the characterisation approach.

Beyond these cases, the decision tree is all about handling some special cases.

In building this framework, we identified five core results which form our remaining conclusions. The first is that some annuity books, and some pension schemes, are quite simply large enough to be self-credible. They can measure basis risk simply using their own experience data and that of the reference population underlying the index that is used for the swap. The bodies using this approach require more than 25,000 lives, a long back history with 8 or more years of experience data and to be stable in terms of socio-economic mix. I will explain why those points are so important in a moment. When this situation applies, our second result is that the M7-M5 model is suitable. There are some special cases though where we suggest an alternative model, which we have called the CAE + cohorts. Again, we will explain that model over the course of our presentation.

We have considered the situation of schemes or annuity companies where the above criteria do not apply, that is, bodies that have less than 25,000 lives or insufficient history or socio-economic stability. The methodology has an approach, which is suited for these cases. It is based upon using external data to enable some proxy modelling. We have called this the characterisation approach. So, the framework we have put together is suitable for all circumstances.

Regardless of the approach that you end up following, be it the characterisation approach or direct modelling using models such as the M7-M5, our fourth result is that a key part of the modelling decision will be the choice of time series. We have highlighted this because it is an oft-overlooked part of the modelling decisions.

Finally, and perhaps most importantly, applying our framework suggests you can get meaningful longevity risk reduction by using index-based hedges.

That leads me to providing evidence for each of the seven conclusions. I will start with the first two that are reiterating the work of the Longevity Basis Risk Working Group and why we are concerned about the knowledge gap and quantifying basis risk.

The left-hand chart in Figure 1 looks back over the last three decades and contrasts the annual rate of improvement in mortality rates of men and women. The top, blue, line relates to men. The bottom line relates to women. There is a gap of about 1/2%–3/4% per annum. This is important because any future divergence between those two genders is something that we can hedge using index-based swaps as there exist published indices, based on the England and Wales population, for men and women.

A similar time period, looking at deprivation quintiles, focussing on England is shown in the right-hand chart of Figure 1. The top lines are the least deprived areas. The bottom lines are the most deprived areas. The scale of the gap between the most and least deprived areas is equivalent to around 3/4% per annum in terms of mortality improvements. This is as big, if not bigger, than the gap currently hedged between men and women. This shows how the potential for different books to be made up of different socio-economic mixes could lead to a problem with basis risk. There is a knowledge gap, and a need to quantify basis risk.
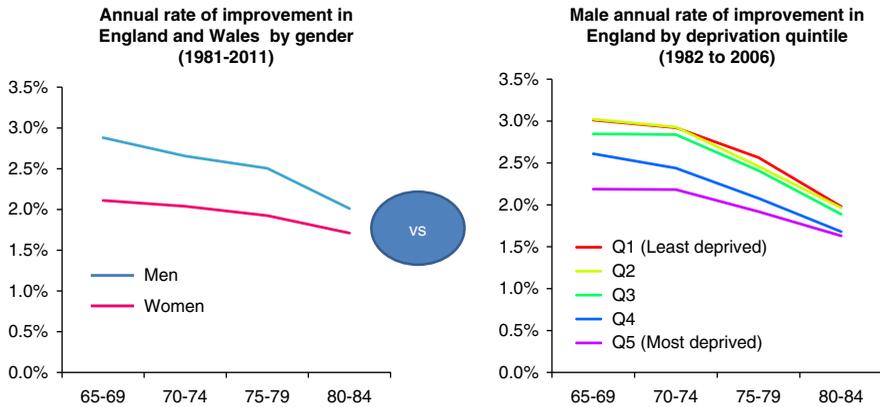
Figure 1. Should we be concerned about basis risk?

Is basis risk something we can already adequately assess or is there a tools gap? Longevity basis risk can be decomposed into three core components: structuring risk, sampling risk and demographic risk.

Structuring risk is related to the fact that we are talking about a specific traded instrument. The challenge here is that any kind of index-based swap will have a series of payoffs, which probably will not match perfectly those of the annuity book or pension scheme. They might be quarterly or annual payments rather than monthly or perhaps more fundamentally the swap might have a tenure of 10 years, whereas the annuity book has payments going out 20 or 30 years.

That is not an issue. We can quite simply simulate the outcomes from a particular contract and from the book and compare them.

The other component, sampling risk, is related to the fact that we would expect the numbers dying by age to have some kind of bell curve shape, but in reality people do not die perfectly in line with actuarial assumptions and we will get sampling noise. Again, this is not an issue and is easy to handle. We can quite simply do random sampling of when the lives and deaths happen.

The issue is we might not accurately reflect what is going on in the book versus the reference population. In particular, the book might be following a different bell curve from the reference population. Further this difference might not be stable over time. This is the element which we call demographic risk. There is a risk that our books might look very different to the reference population and have mortality which changes differently over time. It is this risk which to date there has been no recognised method for assessing.

It is quite an easy task to write down the issues for assessing demographic risk. We have some death rates for the index or reference population by age and by time. We have a similar set of mortality rates but for the book. Our challenge is to model those two sets of mortality rates, but to do so in some kind of simultaneous way that reflects there might be some correlations in the trends between the two groups. If we can answer that modelling problem, then the rest of the issues fall away.

So we have a need to develop a modelling framework to quantify demographic basis risk. It is that framework that forms our decision tree. Mr Villegas and Dr Millossovich discuss this.

464

Mr A.M. Villegas (introducing the paper): For our modelling framework we developed a decision tree which guides users towards an appropriate method for measuring their basis risks (see page 6/7 of user guide). This decision tree has two main branches. One, on the top, is for self-credible books, that is, those books which have enough data to be able to use the data for direct modelling. The other branch, on the bottom, is for other books which might use an indirect modelling approach because of lack of sufficient data.

For an annuity book or pension scheme to be self-credible, we determined that it should have over 25,000 lives, more than 8 years of data and a stable mix. The reasons for each of these requirements are explained below.

Focussing on the case where we have enough data, a sufficient number of lives, we will be able to jointly model the reference population and the book population. We examined survival probabilities from age 60 to 90. We concentrated substantially on this measure, because it is a simple measure that summarises mortality rates and behaves similarly to annuity rates.

With historical data we can fit a model, which may produce something similar to what is shown in Figure 2 with a central forecast but with some uncertainty surrounding that forecast. That uncertainty will come from different sources. First, we will have process risk, which will relate to uncertainty in the time series of the data. On top of that, there will be some parameter uncertainty, that is, the risk of incorrectly estimating the parameters in the model.

Process risk and parameter uncertainty together make up the demographic basis of risk in your hedge when you look at the differences between the two populations. In addition, there will be some sampling risk, which relates to the volatility in the actual mortality experience of the particular book under consideration.

In a book of smaller size, say a book with 10,000 lives as shown at the bottom of Figure 2, the historical observations are much more volatile. If one fits a model that volatility will feed into the fitted demographic quantities. This will cause all the sources of risk to be exaggerated or amplified. First, process risk will be bigger because historic volatility will feed through the projections and the parameter uncertainty will be magnified as there is less data.

There will also be much more sampling risk, and that is a real risk because the uncertainties surrounding the actual times of death of members will be bigger. If we mis-assess the first two sources of risk, namely process risk and parameter uncertainty, it will result in a distorted assessment of the hedge effectiveness and the basis risk.

To illustrate how we arrived at a figure of 25,000 lives for self-credibility, we focussed on a particular point in time and took a cross-section of a fan chart as shown in figure 6.9 of the report.

We considered the difference in survival probabilities for the reference and the book population in, say, 10 years' time, and see how this uncertainty is made up for different sources of risk and for different book sizes.

Looking at the first chart on Figure 3, on the X axis are different book sizes ranging from 5,000 to 100,000 lives. The bars represent the variance in the difference between the survival probabilities
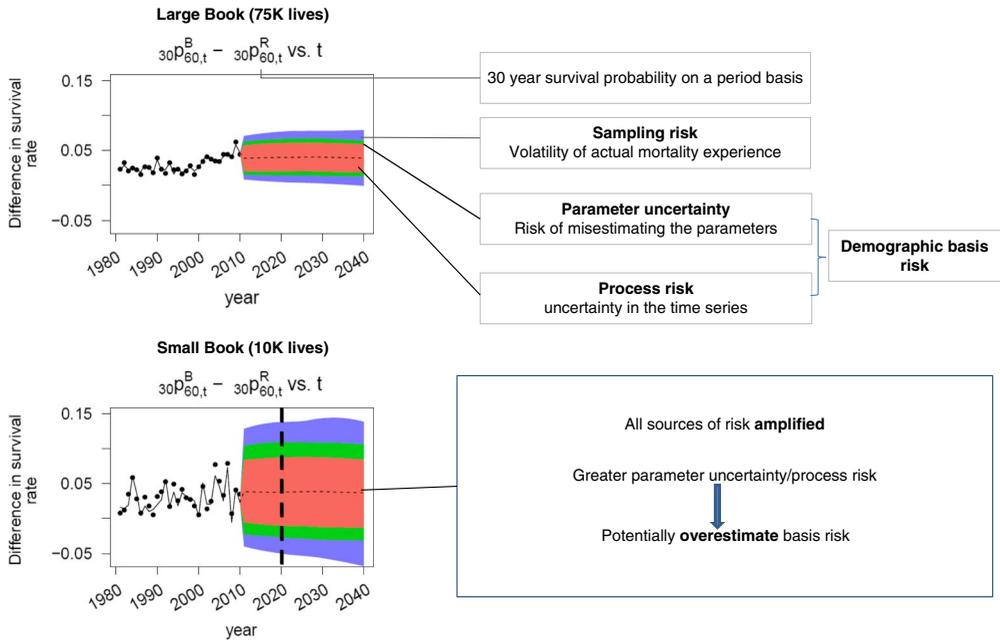
Figure 2. Self-credible? Need 25k lives or more for direct modelling. Decomposition of uncertainty by sources of risk

of the two populations. The colours show what proportions of the totals come from the different sources of risk. As the population size increases, then the variance decreases sharply. At around 20,000 lives the process risk starts to stabilise. Parameter uncertainty is quite large for the smaller books.

We can also consider this type of plot for the book on its own rather than the difference as shown in the second chart on Figure 3. The variance is much bigger, but the volatility, particularly for the sampling risk, starts to stabilise around 20,000 lives. The parameter uncertainty starts to vanish, the sample risk is still bigger and that is a real risk that needs to be assessed.

These types of plot are very informative. Looking at the book is like looking at an unhedged portfolio, and looking at the difference is a proxy for looking at a hedged position after entering into a standardised longevity hedge.

One can assess the effectiveness of entering into a standardised longevity hedge by assessing the decrease between the bars on the centre chart and the bars on the left-hand chart. A calculation of one minus the variance in the difference divided by the variance in the book will give a proxy for the hedge effectiveness. The closer the ratio is to 1, the better the hedge.

The chart on the right-hand side of Figure 3 shows the hedge effectiveness for different books and for different sources of risk. The red line and the reduction between the bars in red, on the centre chart, and the bars in red on the left-hand chart, show the process risk. In the case considered, at around 15,000 lives you notice that there is a pretty sharp decay, which suggests mis-assessing the process risk, and that is exaggerating the basis risk meaning lower hedge effectiveness.

466

Hedged ≡ Difference

$$\mathrm{Var}(_{30}p^B_{60,10} - {}_{30}p^R_{60,10})$$

Unhedged ≡ Book

$$\mathrm{Var}(_{30}p^B_{60,10})$$

Hedge effectivenes
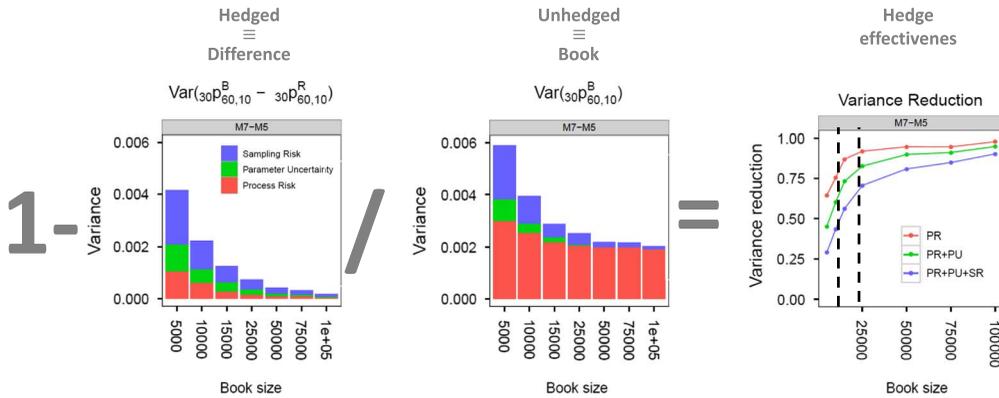
- For book sizes smaller than **15K lives process risk** is unrealistically high, **distorting the assessment of basis risk**
- For book sizes smaller than **25K lives parameter uncertainty is significant, distorting the assessment of basis risk**

Figure 3. Self-credible? Need 25k lives or more for direct modelling. Variance decomposition and hedge effectiveness by book size

The green line shows both the process risk and parameter uncertainty. Exaggerating parameter uncertainty leads to very low hedge effectiveness. Overall, from this analysis we have concluded that around 20,000 lives you have enough data for doing direct modelling.

How did we determine that 8 years of life is needed for direct modelling? The problem here is that for shorter experience there may not be enough data for effective time series modelling. The quality of forecasts is likely to be poor.

The graph on Figure 4 shows two books, one with a lot of data, as shown by the blue line, and one with a shorter data history, as shown by the red line. The latter book might be trapped in some local trends and if we try to forecast it we will have a pretty poor performance.

To assess the situation in a more formal way, we can look at the mean absolute error in our forecast, which is the difference between the actual quantities, the dots in the plot and the forecast, which is shown by the dash lines. We have investigated that for books with different history lengths.

Figure 5 shows mean absolute error plotted against history lengths ranging from 5 years to 20 years. For less than 7 years of data, the mean absolute error, or the quality of the forecast, starts to get really poor. Therefore, we have concluded that you need to have 8 years of data to be able to do direct modelling.

Finally, we require that there is a stable socio-economic mix in the book. If there are changes in the socio-economic mix then these may be confounded with improvements in the mortality rate. In those cases, it is better to use an alternative approach where changes in composition of the book can be directly acknowledged.

In our modelling framework two questions summarise the evidence for self-credibility. The first one is whether you have 25,000 lives and more than 8 years of data, and the second one is whether there have been major changes in the socio-economic mix over time. If the answer to the first question is

467

- For short experiences
  - Not enough data for doing time series modelling
  - The quality of the forecasts is likely to be poor
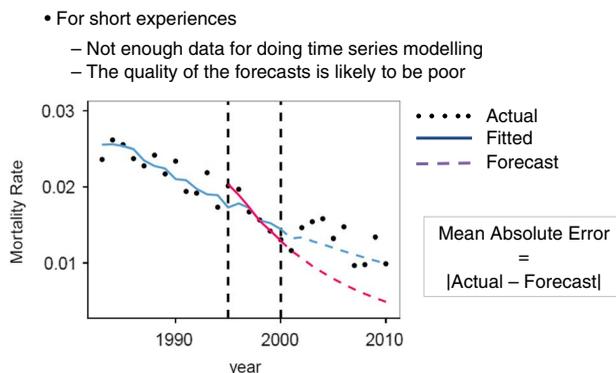
Figure 4. Self-credible? Need at least 8 years for direct modelling

> At least 8 years for direct modelling
> For shorter history lengths, the quality of the forecasts is likely to be poor.
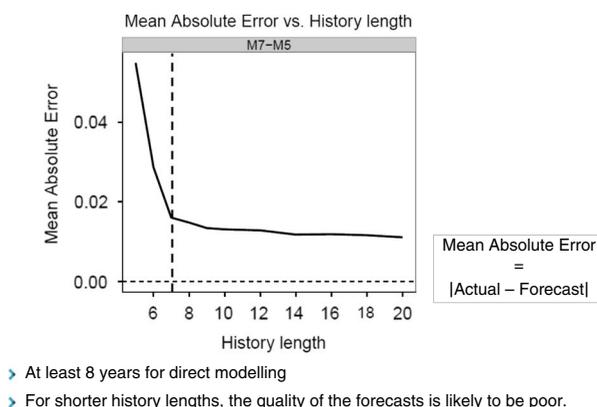
Figure 5. Self-credible? Need at least 8 years for direct modelling. Mean absolute error by history length

yes, and the answer to the second question is no, then you can use direct modelling and in that case you can use the M7-M5 model or in some cases the Common Age effect plus cohorts model.

Figure 6 shows a summary of two population models which we have created after a critical review of the existing literature on these models. Most of the models that are proposed can be classified into three groups. One group is large and uses extensions of the well-known Lee–Carter model. Another group consists of extensions of the also well-known Cairns–Blake–Dowd (CBD) model of mortality. There are some other models that combine features from the two families of models or cannot be classified as being in either of them.

We have critically assessed all these models against a list of criteria that a good and practical model should satisfy. We have divided these criteria into three stages. The first stage refers to criteria that are related to the theoretical characteristics of the model and do not require data to be assessed. We have some criteria that require data and relate to goodness of fit and to the reasonableness of models. The third stage is robustness.

The main highlights of the assessment are set out in detail in the paper. Initially, we consider criteria that do not require data. A very important criterion that has allowed us to filter out a lot of the

468

Common factor
$log \mu_{xt}^i = \alpha_x^i + \beta_x \kappa_t$
Carter and Lee (1992), Li and Lee (2005), Li and Hardy (2011)

Stratified Lee-Carter
$log \mu_{xt} = \alpha_x + \alpha^i + \beta_x \kappa_t$
Butt and Haberman (2009), Debon et al (2011)

Piggyback model
$log \mu_{xt} = \alpha_x + \beta_x \kappa_t + \alpha^i + b^i x$
Currie (2009)

Relative P-Splines
Biatat and Currie (2010)

Three-way Lee-Carter
$log \mu_{xt}^i = \alpha_x^i + \beta_x \lambda^i \kappa_t$
Russolillo et al (2011)

Relative Lee-Carter + Cohort
$log \mu_{xt}^i = \alpha_x + \beta_x^{(1)} \kappa_t + \gamma_{t-x} + \alpha_x^i + \beta_x^{(2)} \kappa_t^i$
Villegas and Haberman (2014)

Multipopulation GLM
Hatzopoulos and Haberman (2013), Ahmadi and Li (2014)

Joint-κ
$log \mu_{xti} = \alpha_{xi} + \beta_{xi} \kappa_t$
Carter and Lee (1992), Li and Hardy (2011), Wilmoth and Valkonen (2001), Delwarde et al (2006)

Lee-Carter (oval)

Other (oval)

Plat + Lee-Carter
Wan et al (2013)

Common Age Effect
$log \mu_{xt}^i = \alpha_x^i + \sum_j \beta_x^{i,j} \kappa_t^{i,j}$
Kleinow (2013)

Augmented-Common Factor
$log \mu_{xt}^i = \alpha_x + \beta_x \kappa_t + \alpha_x^i + \beta_x^i \kappa_t^i$
Li and Lee (2005), Li and Hardy (2011), Hyndman et al (2013))

SAINT model
Jarner and Kryger (2011), Jarner and Moller (2013)

Co-integration approach
Salhi and Loisel (2013)

Cointegrated Lee-Carter
$log \mu_{xt}^i = \alpha_x^i + \beta_x^i \kappa_t^i$
Carter and Lee (1992), Li and Hardy (2011), Yang and Wang (2013)

Bayesian two population
$log \mu_{xt}^i = \alpha_x^i + \kappa_t^i + \gamma_{t-c}^i$
Cairns et al (2011a)

Lee-Carter + VAR/VECM
$log \mu_{xt}^i = \alpha_x^i + \beta_x \kappa_t^i$
Zhou et al (2014)

Gravity model
$log \mu_{xt} = \alpha_x^i + \kappa_t^i + \gamma_{t-c}^i$
Dowd et al (2011)

CBD (oval)

Plat relative model
Plat (2009b)

Two population M7
$logit\, q_{xt}^i = \kappa_t^{(1,i)} + (x - \bar{x}) \kappa_t^{(3,i)} + ((x - \bar{x})^2 - \sigma_x^2) \kappa_t^{(3,i)} + \gamma_{t-c}^i$
Li et al (2014)

Two population M6
$logit\, q_{xt}^i = \kappa_t^{(1,i)} + (x - \bar{x}) \kappa_t^{(2,i)} + \gamma_{t-c}^i$
Li et al (2014)

Two population CBD
$logit\, q_{xt}^i = \kappa_t^{(1,i)} + (x - \bar{x}) \kappa_t^{(2,i)}$
Li et al (2014)

Figure 6. The landscape of two population models. CBD: Cairns–Blake–Dowd; VAR/VECM: Vector Autoregressive/Vector Error Correction Model

models that have been proposed in the literature is whether the model implies a perfect correlation between the book and the reference population. In these models the reference population and the book population move in tandem. So a spike in the reference will be matched by a spike in the book. When you try to forecast with these models you find that the fitted values are very smooth and the uncertainty in your forecast values is going to be very low or non-existent, which will lead you to thinking that you have no basis risk. In contrast, a good model that allows for non-perfect correlation between the two populations should forecast much more uncertainty. This is exemplified in Figure 7.

Some of our other criteria are related to the practicalities of the models. First, for models which comply with the data characteristics, for instance, models that do not require that you have the same amount of data in the reference and in the book population. We have also concentrated on models that are transparent, so that you can understand the assumptions and the meaning of the parameters and explain them to the people that are going to use the model. We have also focussed on models that are not very difficult to implement with a reasonably simple mathematical structure or available software. Using our criteria we narrowed down our list to around nine models that meet the first stage of the assessment.

Dr P. Millossovich is now going to take you through the second stage of our assessment.

Dr P. Millossovich (introducing the paper): The second set of criteria, stage 2, involves goodness of fit and reasonableness of the estimation, particularly in terms of reasonableness of the uncertainty forecast. For this, unlike the previous set of criteria, we need some test data sets and also a common framework for fitting the two population mortality models that we have. Beginning with the data sets

- Models with perfect correlation between the book and the reference **imply no or very low basis risk**

  - e.g

$$\text{logit } q_{xt}^i = \alpha_x^i + \beta_x^i \kappa_t$$

logit $q_{x,t}^i$ vs. t

Book − Reference: $_{30}p_{60,t}^B - {}_{30}p_{60,t}^R$ vs. t
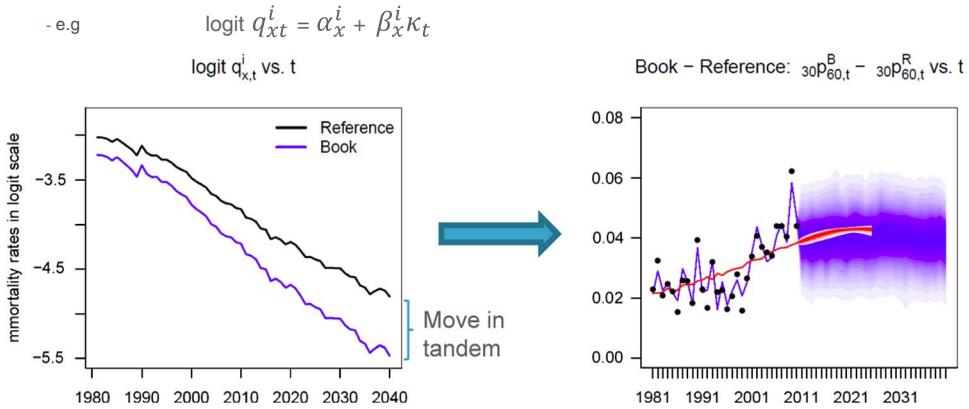
Figure 7. Models with perfect correlation between the book and the reference imply no or very low basis risk

for the reference population, we chose the population of males in England and Wales over a period of 50 years and the range of 30 years of ages, so 60–89.

To create the book our choice is to focus on synthetic data sets constructed in the following way. We have Index of Multiple Deprivation (IMD) specific death rates available at national level. We combine these with typical compositions in terms of the deprivation groups observed in Club Vita pension schemes. In particular, we consider four possible target compositions. The reason we do this instead of working with actual pension schemes is to enable constructing many different instances of pension schemes by changing one of the characteristics like the length, the experience or the size while keeping the other characteristics fixed.

We concentrate on the four shown on the plot in Figure 8 where we have two that have mortalities close to that of England and Wales, or slightly lower, and two where mortality is dramatically different from that of England and Wales. The weighting that we have used is based either on lives or on pension amounts.

As for the common modelling framework, most of the models we analyse can be framed according to two equations:

$$\text{logit } q_{xt}^R = \alpha_x^R + \sum_{j=1}^{N} \beta_x^{(j,R)} \kappa_t^{(j,R)} + \gamma_{t-x}^R$$

$$\text{logit } q_{xt}^B - \text{logit } q_{xt}^R = \alpha_x^B + \sum_{j=1}^{M} \beta_x^{(j,B)} \kappa_t^{(j,B)} + \gamma_{t-x}^B$$

The first equation is specifying the death rate in the reference population. There are several terms appearing on the right-hand side. There is an α term representing the general level of mortality and possibly several time indices reflecting mortality improvements. The β terms are age-specific parameters. Then there is possibly a cohort-specific term. In the second equation, we concentrate on the spread between death rates in the two populations, again on a logit scale. On the right-hand side of the

470

Reference: England and Wales males, 1961-2010, ages 60-89

Book: **Synthetic datasets based on Club Vita data,** 1981 - 2010, ages 60-89
- Constructed by **randomly sampling** from national IMD data
- Splits between deprivation groups as in **Club Vita pension schemes**
- Four different distributions of members by deprivation groups
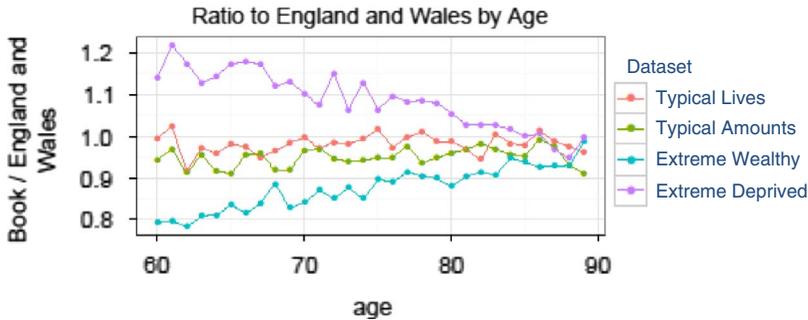
Figure 8. Stage 2: goodness of fit and reasonableness: testing data sets. IMD, Index of Multiple Deprivation

equation the structure is similar. Interpretation, however, is different because the α term now reflects general mortality differences between the two populations, while the time indices, the κ terms, represent improvement differences between the two populations. Again, the β terms are age-modulating parameters. The γ term is possibly attributed to cohort differences between the two populations.

Going back to the two main model families, the Lee–Carter and the CBD, the Lee–Carter family is characterised by having non-parametric β terms, so age is treated as a factor, and the CBD family is characterised by having parametric β terms, so age is treated as a quantitative variable.

Keeping in mind the second equation, the first key finding is that, when we are looking at the spread between the reference and the book, in most of our examples there is not enough data in the book to estimate reliably this non-parametric term. If we insist in doing that estimation, the problem that we might have is exemplified by the two graphs on Figure 9.

In the one on the left we have our estimate of the non-parametric β term versus age, which fluctuates around 0. When we use that term to compute some aggregate metric like the spread between the two 30-year survival probabilities at age 60, the fit is too smooth to capture the variation in the observed differences in the past. Similarly, if we try to use this term to forecast, the resulting confidence bounds are too narrow.

The model behaves as if it implied perfect correlation, or that the β were equal to 0. This means that we should avoid having such non-parametric terms. Instead, we should have either a parametric term or the non-parametric term should be taken from the corresponding equation for the reference population. If we do that then the results are much better both in terms of fitting and forecasting.

If we look at the goodness of fit of the models we are considering, the models should be rich enough to be able to perform well when trying to capture some individual population metrics.

471

$$\text{logit } q_{xt}^B - \text{logit } q_{xt}^R = \alpha_x^B + \sum_{j=1}^{M} \beta_x^{(j,B)} \kappa_t^{(j,B)} + \gamma_{t-x}^B$$

$\beta_x^{(1,B)}$ vs. x

Book − Reference: $_{30}p_{60,t}^B - {_{30}}p_{60,t}^R$ vs. t

- Not enough data in the book to estimate β_x^B
- May result in over-smoothed aggregate demographic metrics
- The model behaves as if it implied perfect correlation
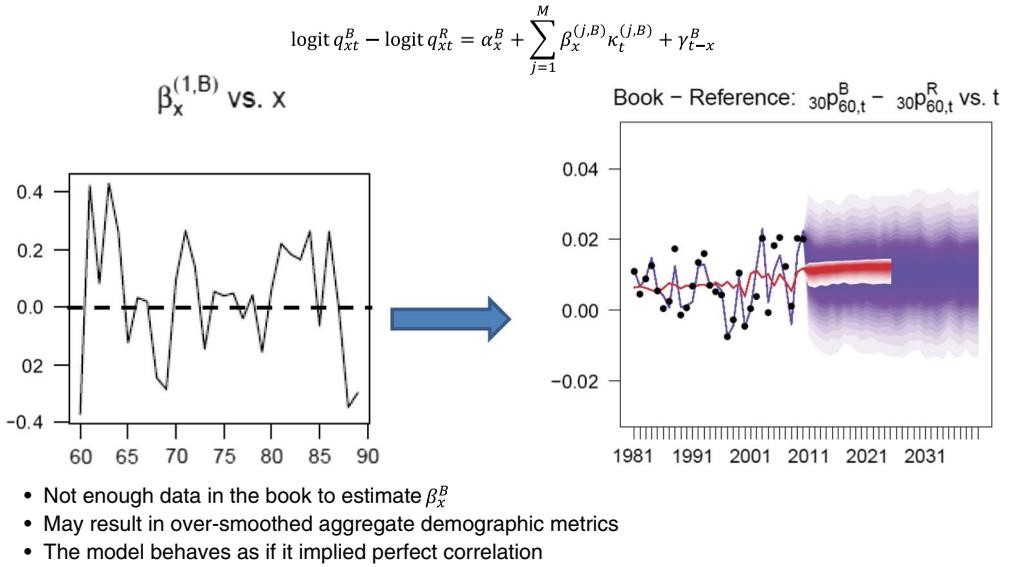
Figure 9. Stage 2: goodness of fit and reasonableness; avoid models with non-parametric book-specific age-modulating parameter

$_{30}p_{60,t}^B$ vs. t
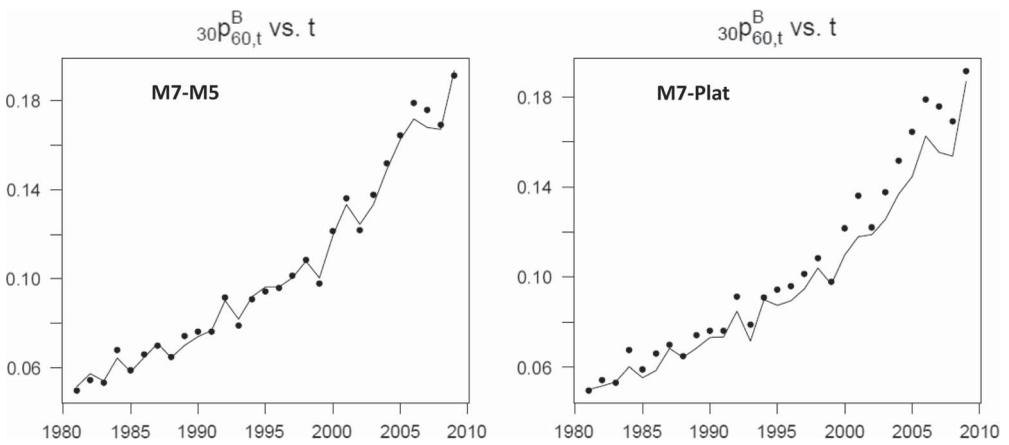
Figure 10. Stage 2: goodness of fit and reasonableness; some model showed poor goodness of fit

In the example shown in Figure 10, we consider the 30-year survival probability at age 60. The left-hand graph shows our preferred choice, the M7-M5. The right-hand graph shows the results from a simpler model, which is not able to reproduce well the pattern of observed probabilities.

If we use an information criterion, that is an index which summarises the goodness of fit accounting for the complexity of the models (penalises models for the number of parameters they have), we can see that the M7-M5 and our second-best model, the common age effect plus cohort, always give the best compromise in terms of goodness of fit and parsimony. These two models dominate other models, in particular those including book-specific cohort effect or a curvature term in the differences between book and reference. Given this analysis, we further narrow down this landscape of models.

472
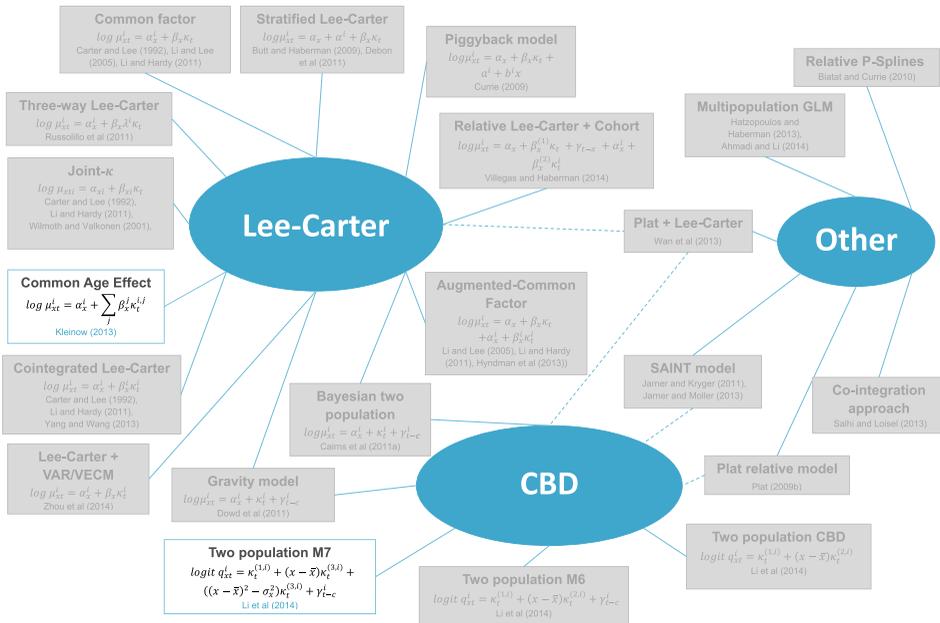
Figure 11. Stage 2: goodness of fit and reasonableness. CBD, Cairns–Blake–Dowd; VAR/VECM, Vector Autoregressive/Vector Error Correction Model
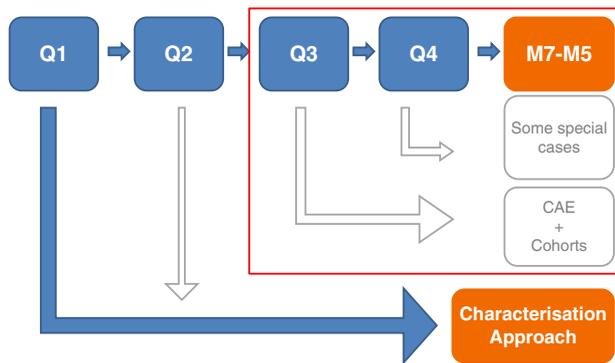
Figure 12. Our modelling framework (stylised)

We end up with the two that are not greyed out on Figure 11: the two population M7, this is what we call M7-M5, and the common age effect.

We are going to focus now on the top-right part of the stylised flow chart shown in Figure 12 to understand when we should use the M7-M5 and when we should choose the alternative common age effect plus cohorts or consider other possible situations.

I will quickly describe these two models and their differences. The M7-M5, starting from the reference population, comes from the CBD family. This means it has non-parametric age-modulating parameters. If we look at death rates in the reference population, for a fixed calendar year on a logit

473

scale, they are quadratic functions of age. The intercept slope and curvature term depends on time, on the κ terms that have to be forecast. Then the last term, the γ, accounts for cohort-specific effect. The spread between the book and the reference, again on a logit scale, for a fixed calendar year is a linear function of age. We did not find it necessary to add additional terms to this expression.

The common age effect with cohorts comes from the other family and is an extension of the well-known Lee–Carter model. It has non-parametric age-modulating terms. For the reference population, that is essentially a Lee–Carter model with an additional cohort term. When you look at the difference between book and reference, again on a logit scale, we again have a similar expression. The α represents level differences, the κ improvement differences and the corresponding β term in front of the κ is taken from the reference population. We do not have a specific β term for the book here.

The description is completed by specifying the model for the time series. Here we follow the main approach in the literature and use a multivariate random walk with drift (MRWD), because we have several time indices for the period terms in the reference population, which implicitly assumes that past historical trends will be repeated in the future. For the cohort effects, we use an integrated process where we take differences first resulting in a stationary process. For the book population, we assume that the κ terms driving differences between the book and the reference follow a stationary process, so will revert to some historical average in the future. A user may want to change these assumptions and use a different time series if he has a different belief.

If we consider the main differences and the relative merits of the M7-M5 with respect to the other models, the M7-M5 is simpler to fit than the common age effect because of the non-parametric terms appearing in the common age effect plus cohorts. On the other hand, in terms of forecasting, the M7-M5 is less satisfactory because five time indices have to be forecast, whereas there are just two in the common age effect plus cohorts. The key difference is in terms of the inter-age correlation structure, which is less restricted for the M7-M5 than for the common age effect plus cohorts. This may be important when constructing or structuring an index-based longevity hedge. This is in case a user wanted to understand what kind of protection is provided by the age structure in the index in respect of the possibly different age structure in the book. If this is a main concern, then the choice should be restricted to the M7-M5. As a final point, it is possible to embed non-base rates in both of the two models.

We expect most users to allow for inter-age mortality correlations. In that case, they should either use the M7-M5, or the common age effect plus cohorts.

If a user wanted to have a book-specific cohort effect, both models can be modified by adding such terms in the specification, although our experience suggests that to do this it is necessary to have a large book size in order to make the estimation reliable.

That concludes the part of the direct modelling where we have shown that for such self-credible books one should generally use the M7-M5 model, or possibly, as a second-best choice, the common age effect plus cohort model.

The indirect modelling is covered by Mr Baxter.

Mr Baxter: Not all of us are fortunate enough to have portfolios suited to the direct modelling approach, but it was very important we spent some time looking at it, because using the indirect approach will leverage everything that Dr Millossovich and Mr Villegas have just told us about.

474
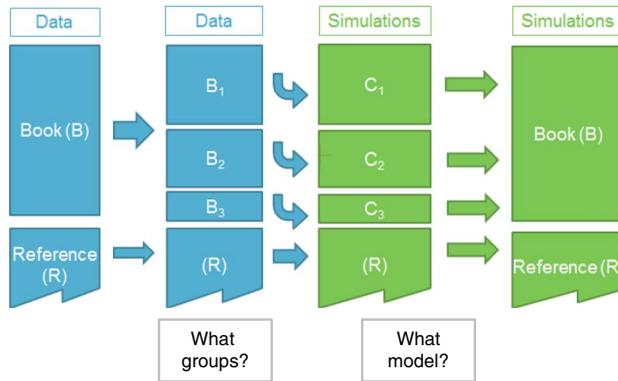
Figure 13. What to do when you have a "small" book?

Simply, for small books of data, if we can find some external data source which we can use to proxy our book, and on which we can apply the methods we have just seen, the problem is solved.

For example, if we have an annuity book or pension scheme, B, as considered in Figure 13 and we want to model this alongside a reference population, R. One thing we could do is segment that book into different sub-groups based upon some socio-economic characteristic. Superficially, that does not get us very far because the sub-groups will still be small. However, if we can find an external data set that we can segment in precisely the same way and apply the methods described earlier to each of the segments, then we can derive some simulations for the different socio-economic groups, C1, C2 and C3, on the schematic. We need not have three groups. I have just chosen three for the purposes of illustration.

Having derived our simulations, it is a relatively easy to weight those simulations in proportion to the exposure we have within our specific book to get a simulation for the book as a whole.

This poses several challenges in determining what we should use for the external data set, how we should group it and what modelling we should do.

If we consider the first of these questions, whatever data set we use has to have a characteristic in it, which we can also use to segment our book. Notwithstanding that, there are a number of other criteria. Within the detailed paper we have highlighted a few potential data sources. You could use the publicly available Office of National Statistics (ONS) data, segmented by postcode-based deprivation. You could look to the Continuous Mortality Investigation (CMI) data or perhaps the Self-Administered Pension Scheme investigation data segmented by pension amount. There are other third-party data sets, such as the Club Vita data set, which would allow you to segment by more than one dimension of socio-economics.

In the paper, we give an example of creating these groups for both the ONS data set, shown on the top-right of Figure 14, and the Club Vita data, on the bottom-right of the Figure 14, which uses two dimensions, a pension dimension and a postcode-based deprivation dimension.

In forming those groups, there are certain things that need to be taken into account. We suggest a series of principles in the paper. Alongside those principles, we indicate statistical methods you can

475
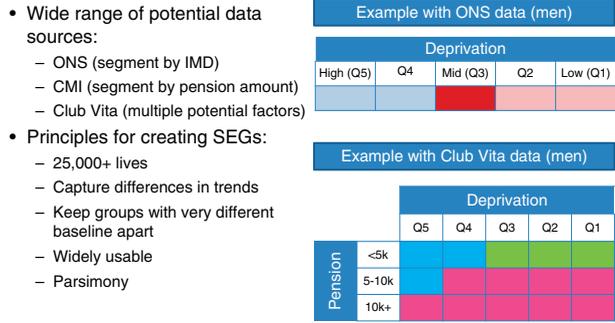
Figure 14. Characterisation population and groups. ONS, Office of National Statistics; IMD, Index of Multiple Deprivation; CMI, Continuous Mortality Investigation
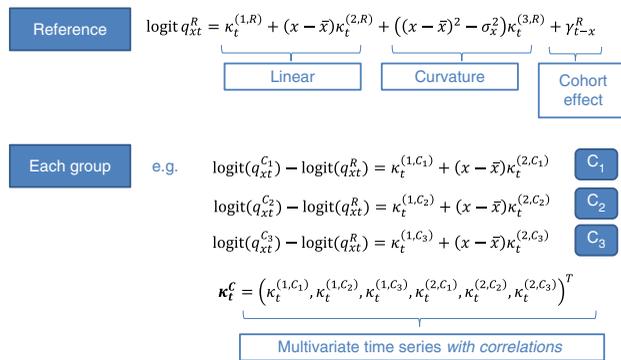
Figure 15. A multi-population M7-M5 model

use to achieve groupings consistent with those principles. I will not dwell on what those principles are; suffice to say that the groups you get out of that process have to be large enough that you can apply the kind of modelling techniques Dr Millossovich and Mr Villegas described, that is you need to have at least 25,000 lives in each. You want the groups you create to be widely usable, and capture any differences that there are in improvement trends as you are using them as a proxy to the underlying trends that drive demographic risk.

Once you have created those groups, you then need to be able to model the subsets of this third-party database or data set alongside the reference population. It is relatively easy to take the model that Dr Millossovich and Mr Villegas described earlier and extend it. In this situation you are almost certainly going to answer "yes" to all the questions in our decision tree and so use a multiple sub-population extension of the M7-M5 model (Figure 15).

The model has linear, curvature and cohort effect terms for the reference population. Each of the sub-populations that we are looking at, C1, C2 and C3, in Figure 16 has its own M5 equation, the linear expressions on the right-hand side of the equations.

We get some extra complexity at this stage because we have three different populations. The sub-population equations have six different time series terms, κs, that we need to model. We do not want

476
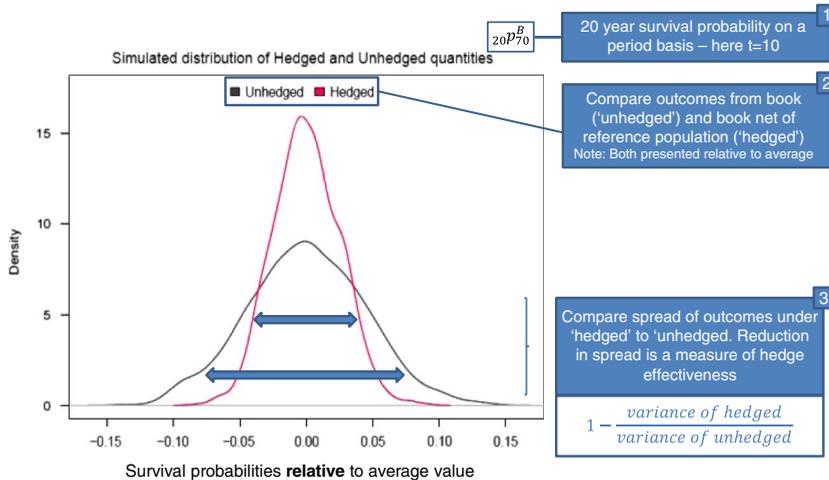
Figure 16. A simple measure of hedge effectiveness

to model each of the three populations in isolation with the reference population. Instead, we want to model them simultaneously with the reference population. This means we need to take into account the fact that there may be some correlations. You could imagine a certain socio-demographic group diverging from the reference population in one direction, whilst others go in the opposite direction to average this out.

We have to give some thought to the correlation structures . This adds a small amount of complexity to the models for the sub-populations, but the trade-off for that complexity is that once you have done it once, you have a set of simulations you can reuse over and over with different books.

I hope that gives you a good flavour of how we can use the techniques we described earlier, even when we have a small book.

Moving on to what happens when you apply this methodology in practice. We are not going to focus on showing you the detailed results of the actual model fit. They are all in our supporting paper. Instead, we are going to focus on the results of using the methodology for a selection of different books.

We selected five books, all of which are based upon real pension scheme data. We have three large books; A, B and C. These are each big enough to facilitate direct modelling. They have more than 25,000 lives and more than 8 years of back history. A great starting point to compare and contrast direct modelling with the characterisation approach. We also have a small book, one where you would rely on the characterisation approach. Finally, we also have a medium-sized book with 20,000 lives. This is in a slightly grey area where it may be possible to use direct modelling.

To give you a flavour of the socio-economics of each of those books, on the right-hand side of the table on Figure 17 we have pie charts giving the split by the ONS IMD groupings that I showed you earlier. We do not need to dwell on the charts, other than to draw your attention to large scheme C, where there is a sizeable grey area in the pie chart. Over a third of that particular book could not be mapped to an English IMD. That will be important when we move on to the next stage.

Annuity book	Annual exposure[1]	Exposure period	IMD split **Low** **Mid** **High** Unknown
Large A	28k	19 years	
Large B	28k	13 years	
Large C	28k	10 years	
Medium	20k	10 years	
Small	12k	19 years	

Notes:
1. Exposure in final year of data

Figure 17. Testing the approach

Before I show the results, I should show you what the results pertain to.

Figure 16 shows a cross-section through a funnel of doubt ten years into the future. We are looking here at 20-year survival probabilities from age 70 and at the position before and after hedging. The pink and black bell curves on the chart are measuring the survival probability relative to their respective average value. The black curve shows the unhedged position; the book if we did nothing. The hedged position the pink curve, is the difference in the results between the book and the reference population. This is the residual risk after implementing an index-based swap.

We are going to use a simple measure – the variance of the bell curves, and specifically how much is it reduced by moving from the unhedged to the hedged position.

Figure 18 shows the results of our five books using that broad indicator of hedge effectiveness. The first column of numbers relates to the direct modelling. The other three columns of numbers show different variations of the characterisation approach.

The first two columns of numbers allow us to compare and contrast the direct modelling and the characterisation approach. The broad similarity in the numbers, large scheme C aside, shows that the characterisation approach is a credible alternative to direct modelling, giving a reduction in variance of a similar order of magnitude. Furthermore, if we focus on large A, large B and the medium scheme, we can see that there is a modest difference in the numbers. We refer to this as the residual basis risk. It shows that there is a possibility that the characterisation approach may not necessarily capture all of the basis risk that might be there. We are proxying the book by some socio-economic groups. It could be that there are some other aspects of the demographic risk specific to the book in question which those groups have not quite fully captured. The differences are typically quite small. Indeed, we see with large scheme B it can work the other way. You might find lower hedge effectiveness when you apply the characterisation approach. That shows the other side of the problem, digging beneath the surface to look at socio-economic groups and what is going on therein might add additional perspective on the modelling of basis risk for your book. This suggests you might want to use the characterisation approach from a model risk perspective even when you are following the direct modelling approach.

Large C seems to be a bit of an outlier. That was not a surprise. We did not know reliably how to map over a third of that book to the characterisation approach. We had to make an educated guess for that third. That the results are an outlier tells us that this is not a great thing to do. You want to

478

Annuity Book	Direct Modelling M7-M5 (VAR with Constant)	Different characterisation approaches		
		Club Vita Characterisation (VAR with Constant)	ONS Characterisation (MRWD)	ONS Characterisation (VAR around Trend)
Large A	78%	84%	77%	88%
Large B	80%	79%	73%	85%
Large C	65%	77%	73%	84%
Medium	77%	80%	75%	85%
Small		75%	70%	79%

> Characterisation approach is a **credible alternative** to direct modelling
> **Modest residual basis risk**
> Need to be able to reliably map book onto characterising groups
> Results depend on dataset used
> Results depend on how interpret signals in the historic data

Figure 18. Example hedge effectiveness results. ONS, Office of National Statistics; MRWD, multivariate random walk with drift; VAR, vector autoregressive

make sure that, however, you are doing the characterisation approach, you pick a data set that is reliable, and that can be used to map most, if not all, of your book across to.

The second and the fourth column of numbers are applying the characterisation approach. The difference here is the data set used to calibrate that approach. Different data sets will give you different perspectives on trends by socio-economic groups and slightly different answers. You need to give some thought to what is the most relevant data set to use.

Finally, note there is quite a bit of difference between the numbers in the last two columns. Both use the ONS data set to derive our characterisation. The difference is in the time series that we have used. The figures in the second to last column are derived using the MRWD. The figures in the last column are derived using the vector autoregressive around trend (VAR around trend).

The MRWD allows the difference between the book and reference population to grow with unbounded variability. Vector autoregressive processes have a bounded element to the divergence. This limits how far away from each other the book and reference population can move, which affects our assessment of basis risk.

It is therefore important to think carefully about what is the appropriate time series to use, whichever approach we are following, be it direct modelling or characterisation. We need to take care in how we interpret the signals in the historical data and understand the implicit assumptions in whatever model we then build around that.

Further, the table demonstrates meaningful longevity risk reduction is possible from index-based hedges; with 70%–80% of the variability being removed. This encourages us that our work can help stimulate people's modelling and, hopefully, the market.

In applying this work there are some practical challenges to be aware of. We highlight a few of these in sections 9 and 12 of our main paper. There are a couple worth drawing to your attention.

Firstly, all of the work we have described has been implicitly assuming that either you are focussing on a single gender, or in some way unisex lives, to attain the 25,000 lives. In reality, every annuity book and pension scheme is likely to have men and women in it, and you are likely to be building hedges taking that into account. This is perfectly manageable within our framework, but you need a few extra correlations, potentially between the two reference populations, and potentially also between the book populations.

All the modelling we have presented thus far has relied on past data. That puts us at risk of the accusation of "driving paying too much attention to the rear view mirror". We would argue that you have to use the past data to guide and inform your modelling process, but there is an important role for the use of judgement in gaining a broader understanding of what might be driving longevity trends. We certainly see a role for that judgement in the choice of time series.

Finally, it is very important when using the characterisation approach, particularly for smaller books, to think carefully about what data you have available and what data you are going to use to proxy your book. There needs to be a marrying up, both in terms of the available data, but also in terms of it having a consistent meaning to ensure you are using the data appropriately.

Where might this work go next? We are conscious that we have left a few questions open for future research. We would love to pick up the mantle on the time series part to see if we can come up with some kind of decision tree that helps users through that part of the decision-making process.

Another area that we think is important, and is a key aspect for consideration under phase 2 of the research, is the development of hedge effectiveness metrics. We have used some simplistic metrics to date to illustrate the methodology and to give comfort that we are getting good and meaningful insights. That does not mean that those are the right metrics to use. We would like to explore what the most appropriate metrics are.

We are also conscious that there is an issue around making regulators comfortable with the methodology and helping users through the regulatory discussion and evidential process.

Summarising our key conclusions:

  i. there is a need to quantify basis risk;
 ii. we believe there is a need for a modelling framework and that we have delivered one;
iii. within that framework, we feel that some data books are self-credible and can rely on their own data.
 iv. For those books, we have suggested the M7-M5 model as a suitable starting point for modelling. Notwithstanding this, there are some exceptions where you might use other models, in particular, the CAE + cohorts;
  v. for smaller books we have also developed an alternative indirect approach, which we have named the characterisation approach.
 vi. the choice of time series used is a key decision.
vii. the framework is suggesting index-based swaps can offer material risk reduction.

480

We hope this work makes assessing basis risk accessible to everyone in the industry.

The Chairman: We now open the paper and presentation to the floor for discussion and questions. After the discussion, Mr Gaches will sum up the discussion from the panel's perspective. Firstly, I will ask Professor Andrew Cairns to open the discussion.

Prof A.J.G. Cairns, F.F.A. (opening the discussion): First, I will make some general remarks. In terms of the things that I liked in the paper, there was a very careful step by step guide: the four questions that users might go through in terms of choosing a model, a particular approach that they might use and reference to the size of the book population, which is clearly a very important point.

Another highlight is the list of desirable criteria. It is something I have contributed to over the years, but the list of the most desirable criteria is getting longer and longer as the years go by. No model can ever satisfy all of these.

In section 6.1 of the report, some or all of the detail of criteria 4, 12 and 16–19 are new. I did not understand the point some were trying to address and elaboration on those would be helpful.

In the numerical results there are some graphics where the uncertainty and the hedge effectiveness, and so on, are divided into process risk, parameter uncertainty and sampling. That is a nice sub-division and it highlights the issue that when you are doing modelling work, you should not just get your best estimate set of parameters, but also consider the impact of sampling risk and the parameter uncertainty. The key element of that is the uncertainty that you get in the drift in the random walk, which is very important the more you go into the future.

Figure 6.1 in the report is where the authors talk about the convergence of mortality rates at higher ages. You have quite big differences at younger ages and rather small differences at higher ages. I wonder, when you look at the correlation between the different sub-groups, whether you find higher correlations between people in different groups who are aged 85–90 compared with people who are, say, aged round about age 60?

There are a couple of places where the authors and the users could do things slightly differently. This is in reference to the M7-M5 model and models that go down that route. For simplicity, possibly because of the correlation matrices, the authors proposed that the difference between the book population and the reference population is independent of the dynamics of the reference population. That is quite a strong assumption. If you assume that they are independent, then, when you are actually modelling the book population itself, it is going to be the sum of two independent risks. Therefore, the risk associated with a book population is going to be larger, but also if you do not model correlation between the difference on the one hand and the reference population on the other hand then potentially you might end up over-estimating how much basis risk there is. If you model correlation, which should be fairly straightforward to do, with only a few more parameters to estimate, you would get improved but also higher estimates of hedge effectiveness as a result.

A point where I disagree with the authors is the use of Bayesian methods. Halfway through the paper, when they are looking at various models, there is one model that is pushed out of the frame. The fact that the model goes out of the frame does not matter, but the reason given was the use of Bayesian methods. I agree the Bayesian methods are more difficult to implement, and there is

481

Population k, year t, age x

$$\log m^{(k)}(t,x) = \beta^{(k)}(x) + \kappa_1^{(k)}(t) + \kappa_2^{(k)}(t)(x - \bar{x})$$

(Extended CBD with a non-parametric base table, $\beta^{(k)}(x)$)

- 10 groups, $k = 1, \ldots, 10$ (low to high affluence)

- 21 years, $t = 1985, \ldots, 2005$

- 40 ages, $x = 55, \ldots, 94$

Figure 19. Modelling the death rates, $m_k(t, x)$. CBD, Cairns–Blake–Dowd

a lot more programming work to do. But there are significant advantages in using Bayesian methods including the ability to tie time series estimation into the stage 1 estimation process, the ability to incorporate missing data, the ability to mitigate parameter estimation bias when you are dealing with small populations and output that immediately gives you the ability to assess parameter uncertainty.

If you wanted to go down the M7-M5 route, the paper is a little bit short on methods for dealing with smaller populations below the 25,000 threshold.

I would promote, therefore, the use of Bayesian methods. Bayesian methods are very much designed to tackle the issue of small population sampling risk as well as the other parameter estimation risks.

To move on from comments on the paper, I wanted to draw attention to some work that I have been doing with Blake, Dowd, Kallestrup-Lamb and Rosenberg, that complements tonight's paper.

We have been using Danish register data, which has high-quality data on all residents, going from the early 1980s onwards. This data set allows us to construct something similar to the data set that we have been talking about but in perhaps a more comprehensive way, because of the availability of many different covariates.

Two of the covariates that we have focussed on are net wealth and gross income and how you use those as predictors to sub-divide the population. As people are well aware from an analysis of various data sets, including pensions data and CMI-type data, if a person has high income or high wealth, then you can be pretty sure that the person is indeed affluent and they will be, in general, healthy. The converse is not true. If a person has low income or low wealth that does not necessarily mean that they are not affluent in some sense. The data that we had available to us allowed us to do a lot of work on this issue. Eventually, we just developed a very simple combination which we called "affluence", which is wealth +15 times income. What we finally found was that low affluence did indeed predict poor mortality, which is not the case if you just look at income on its own or wealth on its own (Figure 19).

Here is the model that we have looking at. It is similar to the one that Mr Baxter just discussed with the exception of an additional non-parametric age effect (Figure 20).

The results for the Danish data strongly support what we have seen in the results for the UK data. Ten populations have been sub-divided, all of equal size. At the bottom we have the most affluent people, and up at the top, the black line is the people who are least affluent. You can see here what the results of that
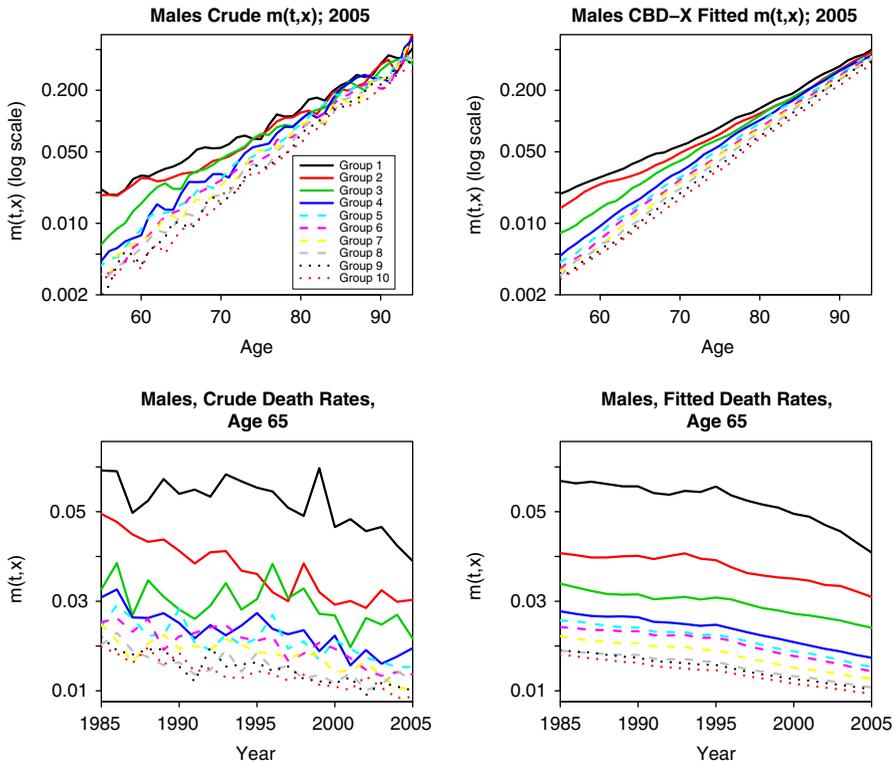
Figure 20. Model-inferred underlying death rates 2005. CBD, Cairns–Blake–Dowd

fitting are. Even though the original data comes from a relatively small population, particularly when you divide it into ten sub-groups, the modelling smooths out all of the sampling noise without actually losing the essential characteristics of that data set. What you also see is convergence at higher ages. One of the other things we have seen in this data set is a slight divergence between the most affluent and the least affluent. Our parallel approach complements what the authors have been doing. Perhaps, one difference is that we are treating all of the ten sub-populations as carrying equal weight. We are not trying to model the national population first and then model the sub-populations.

Mr S. Rimmer: I want to ask a question about the choice of the time series model that models the difference between the reference population and the book population. You had chosen something that is first-order autoregressive. That means the long-term average of the difference between the reference population and the book population, by construction, stays constant. When we are trying to determine the effectiveness of a hedge between the reference population and the book population that seems quite a strong assumption to make. Is not part of the uncertainty that the long-run average of the difference between those two groups will not be the same? Have you not almost reduced the uncertainty by the construction of the model?

Mr D.A. Shaffer, F.I.A.: My question follows on from the last one. In figure 69 you showed of hedge ratios, hedge efficiency, without focussing on the detailed calculations, the numbers were much higher than I was expecting, particularly for the very small books of business. In fact, there did not

seem to be very much variation. Perhaps, that was because of the different modelling approaches that you had for small, medium and large books.

Intuitively, I would have expected that for small books, a population hedge would not be terribly useful. Conversely, a hedge would be useful for a large book. I would be interested in your comments on that.

Mr J. Lu, F.I.A.: According to your decision tree, the belief about whether the book has a specific cohort effect is important. Firstly, there is a belief that the cohort effect is highly correlated with smoking cessation, and smoking cessation itself is correlated with socio-economic groups. Mr Baxter used to say that there may be a cascading effect where people in a higher socio-economic group gave up smoking first, to be followed by people in lower socio-economic groups. That potentially implied differences in the timing of cohort effects for different sub-populations. I wonder if you would comment on that. Secondly, before 2009 or 2010 the industry used to think that the middle of the population cohort was around people born in 1926. After that, we believed it is actually centred about 1935. This change in understanding had financial impacts. Because of the theoretical as well as financial impact, the issue of cohort effect is important. Could you elaborate on the sensitivity to these assumptions as well as the certainty that there are no differences in cohort generations between pensions schemes?

Mr R.S. Fitzgibbon, F.I.A.: A question on the choice of data you used for the reference population. I think it was from 1961. In the most recent CMI working paper, there was some doubt cast on the mortality experience from the 1960s, and so the data used in the most recent CMI core projection model starts from 1975. To what extent are your results sensitive to your choice of data?

The Chairman: I will ask the panel to respond to those questions.

First of all, can we consider Mr Rimmer's question about time series and the implied assumption that the long-term difference between the reference and the book is constant, and whether or not this assumption is too strong.

Mr A.T. Gaches, F.I.A. (responding): We agree that the choice of time series is an important decision. Our findings suggest that hedge effectiveness is sensitive to that underlying assumption. We saw in figure 18 that when using the characterisation approach, the choice of time series could in broad terms change hedge effectiveness from around 80% to around 70%. In our main report, section 9.2.3 provides some further commentary on that issue.

Yes, the approaches we have adopted are typically MRWD for the reference population, and vector autoregressive for the book, relative to the reference. We have taken those as the initial starting point, because they are what have traditionally been used in the established literature. We are aware that other approaches could be used. Some of these avoid that assumption of bounded variance, the one that you particularly, and rightly, highlighted.

I would just echo some of Mr Baxter's comments, that the broader consideration of the choice of time series is expected to form part of phase 2 of this work. For now we hope that we have highlighted that this is an important issue. This phase does not provide all the answers to all the questions that we might want to ask.

Mr Baxter (responding): Mr Shaffer's question related to an impression of quite high hedge effectiveness statistics for smaller books. There are two things to note. As the book size drops, so the

484

sampling risk grows and the hedge effectiveness statistics will come down accordingly. Nevertheless, you would still expect a reasonable level of hedge effectiveness for smaller books such as our example with 12,000 lives, because trends which affect the population as a whole will, to some extent, pass through to the book. Perhaps the key question is what happens when we move away from the relatively simple hedge effectiveness statistics used here to consider other aspects of the variability. We could start asking questions such as: what is happening to the 1-in-200 event, which is particularly important for capital reserving. That is something which we can readily calculate. We just picked a simple statistic for today's purposes.

Mr Villegas (responding): Regarding Professor Cairn's comments, there is lots of research that has been done on single population models, which we have expanded. What we have presented is a short description but we are working on something which requires more detail.

In relation to the use of the Bayesian method, it is a good point that these methods are particularly useful when you have small books. That will leverage your data and make your estimation probably less uncertain. However, we have found that fitting this type of model is typically a time-consuming approach when you do it the first time and it might be hard to understand for some people. We recognise you may want to consider the Bayesian method if you want to use your data with smaller sample sizes.

There was another point regarding the use of correlations between the reference population and the book population. It is true that that to some extent we are over-estimating the basis risk, because we are ignoring any correlation. However, if you have a limited time series, say a book of around 10 years of time and then 30 years of time in the reference population, then you run the risk of having a mis-estimation of this correlation.

The Chairman: There was a question around the cohort effect in the book, and the link to smoking cessation and the subsequent link to socio-economic groups.

Mr Gaches (responding): I agree about the benefit of capturing book-specific cohort effects, if possible. The challenge is whether it is possible. We have found that even using relatively large books of the size which practitioners would be able to use for direct modelling, there simply does not seem enough data to demonstrate the differences in cohort in the book relative to the reference population. There are two occasions where we might be able to look at book-specific cohorts. One is with very big data sets. So if you have a big enough data set and can see an effect, it should be allowed. We would be very keen that practitioners did that. Considering some form of residual plots may be a way of checking for that. The other way of dealing with this is to consider whether the book in question is particularly likely to have a different cohort effect. You raised smoking cessation. Suppose we were looking at a pension scheme for a tobacco manufacturer. That might be the kind of book where you might expect a different timing of smoking cessation. It might be in circumstances like that, that you would want to make allowance for a book specific cohort effect.

A final comment is that if you were to use the characterisation approach with a very large characterising population, split in some way between high, medium and low, that might be a way of getting three very large data sets. In that kind of situation you may be able to see differences and apply that through to smaller schemes in terms of timing of cohort effects. However, because of the amount of data needed, maybe you would want to be looking at a parametric adjustment to capture a book-specific cohort effect.

Mr Villegas (responding): In our report we set out some additional work on measuring the parameter uncertainty when you try to estimate book-specific cohort effect. Once you allow for that, the uncertainty is huge and it is so unreasonable that you would not want to use it as a starting point. That is why we advise that if you believe in a book cohort effect; it has to be based on your knowledge of the data, on the socio-economic conditions of the book rather than what you have observed from your mortality data because it is probably going to be very hard to pick it up.

The Chairman: There was the question from Mr Fitzgibbon about the choice of data from 1961, and whether there are issues in the 1960s data. I do not know if anyone here is an expert on ONS data with thoughts on that?

Mr Villegas (responding): I am not an expert on the data but I can comment on whether we would have reached different conclusions if we had used a different data set for England and Wales. When assessing the hedge effectiveness, what is most important is the difference between the two populations. In some sense we are considering the same data set for the reference and the book population. There may be different forecasts of the trend, but the most likely thing that is going to occur is the assessment of hedge effectiveness is going to look quite similar, because that is what your index hedge is providing you with protection for. Probably, the models that we would have picked would be pretty much the same.

Dr M. Bajekal: This is a comment from a non-actuarial perspective, an epidemiologist's perspective. It seems to me that the basis risk that you are looking at and trying to quantify is based mainly on socio-economic differentials. But in epidemiology, when we compare the performance of different hospitals, say, we not only control for age and sex and the socio-economic risk, but also what we call the case mix or differential burden of disease risk between groups. How sick are the patients coming into hospital? I am sure in the north of the country there will be a different profile of healthiness even if the socio-economic distribution of patients is very similar to a hospital in the south. We find is that mortality rates for socio-economic group 5, or the most deprived, is very different (higher) in the north of the country than it is in the south of the country. There is much more regional variability in health outcomes in the lowest end of the deprivation distribution than there is in the very affluent end. These within-deprivation mortality differences may partly be due to differences in case mix. What is your view on this?

Mr Gaches (responding): It is worth looking at the two different forms of modelling. The direct modelling does not specifically do any segmentation. Where direct modelling is applied, it is all about looking at what the data says so in that sense there is not the segmentation which may or may not capture the differences you want.

It is more of an issue for the characterisation approach. Certainly, in terms of defining the characterisation approach, one of the key issues is about balancing complexity of the groups, what is available in the data and what is really going to matter.

If we are looking across the whole of the UK population, then the differences you have described are certainly a shortcoming of using IMD without some kind of regional overlay. If we are looking at the application to insurance books, and pension schemes, and we consider the kind of member who typically would have the benefits of those, what we find is those books are rather underweight in individuals from the lowest socio-economic groups such as IMD 5 and the more deprived areas. Some of the factors which go into the IMD are things like areas having long-term unemployed and

486

high rates of long-term sickness. The kind of individuals involved will often not have had the chance to build up the benefits that we are talking about. I fully accept the need to capture the factors which are important. In this particular application this may be one where those differences within the lowest IMD groups have less impact in practice.

Prof D. Blake: My question relates to the regulator. With all this great work that you have done if you want to get the regulator to give capital relief for those using index hedges, there should be a parallel campaign to work on the regulator at Solvency II level. For example, to engage them in understanding the results concerning the degree of hedge effectiveness. This may be more difficult than you think.

The research that Professor Cairns and I have done also shows the effectiveness of index hedges can be very high, at around 85%. But this is well below the 99% level that the regulators expect to see before they will give capital relief. You need to persuade the regulator that 85% effectiveness for an index-based longevity hedge is "equivalent to" the 99% effectiveness of other types of hedges.

In short, if you want this study to be useful for the industry, you have got to try to get the regulator engaged early on and to persuade them to offer the level of capital relief that will allow the index hedges to work and to be effective.

Mr Baxter (responding): We agree with the importance of engaging with the regulator on this issue. We were invited last Friday to speak to the Prudential Regulatory Authority (PRA) who asked us to do an in-house presentation of the material. The regulator is certainly looking at our work and, we would hope, has an open mind to seeing it applied in practice.

Dr P.J.K. Sagoo: Thanks Professor Blake, that is a very good point. One of the early mission statements of the LLMA was really to get work like this in front of a regulator. We had about 16 people in the audience from the PRA on Friday. The Dutch regulator similarly has expressed interest in also keeping in touch with the work that is going on and will factor it in to what they are seeing from firms as well.

In the Dutch market there have been three transactions. All of those have been presented to the regulator with a basis risk analysis, T.

The Chairman: Staying with the theme of making this practical, has anyone any thoughts on what the potential phase 2 should be looking at, and what we should be focussing on to turn this into something usable and practical?

Mr M. Ashmore, F.I.A.: A next phase should be a cost-benefit analysis. At the minute people have a choice. They can enter into a full indemnity longevity swap or they can enter into an index-based swap. You have now provided a mechanism for quantifying the basis risk that would remain if they did an index-based swap. It is natural then to consider the cost and benefits of doing one versus the other taking into account the capital needed to support the basis risk?

Mr Lu: Following on, it would be useful to back-test some real-life examples. For example, one could investigate the outcome of a pension scheme of a certain number of people that chose to hedge a longevity index-linked instrument 10 years ago. It will provide a more tangible way for us to understand how it can work in practice.

Mr A.J. Jeffery, F.I.A.: I work for the PRA. These are my comments, not theirs. I was not at the PRA meeting last week, so I do not know what anybody else said.

I do not find the hedge effectiveness metric very useful in a couple of directions. Firstly, in determining what that means financially. If I was a general risk manager who knew nothing about this, I would want to know how much this could be incorrect financially. I would like to get a feel for how stable that is. Is this the difference between two distributions or is it the ratio? What does the residual distribution look like? The other thing I would want to know is what are the implicit assumptions and what could go wrong with them? I have heard some of that conversation. I do not know enough about the mathematics to be able to deal with that myself but I heard some interesting comments bringing those points out. It would be helpful if you can give a rough feel for the real risks. That will help you make a much more credible case in terms of communicating to people outside the very narrow longevity experts.

The Chairman: Does anyone on the panel want to comment about the hedge effectiveness metrics?

Mr Gaches (responding): The illustrations we provide in the presentation and in the report focus on survival probabilities as a metric. This simple metric is designed to capture some of the characteristics of annuity rates in the sense that it looks at mortality rates over a range of ages. The analysis underlying the selection of models also looked at survival probabilities at specific ages. In terms of the actual hedge effectiveness measure, which is probably the main focus of your point, you are right, we focussed throughout on percentage reduction in variance. This is a simple, well-established, often-quoted measure. That is not to say that other measures, such as the impact on tails of distribution, are not of more relevant practical interest to those considering longevity basis risk.

Investigation of a wider range of metrics, annuities, life expectancies, and maybe other metrics, as well as a wider range of risk measures would be expected to fall within the intended phase 2 of the project. When the request for proposals came out, these investigations were specifically in phase 2 rather than phase 1.

The final thing that I would point out is considering alternative metrics and measures is really just a case of processing the results of the work that has already been done and pulling out some different numbers and presenting results in a different way. It is a relatively easy step to do on the back of work that has been done to date. I fully accept that a wider exploration for a range of metrics and measures is critically important.

Prof Cairns: Still on the topic of the metrics, the last few comments have been focussed on the regulator. But equally there is a challenge in terms of persuading pensions scheme trustees or life insurance boards of directors that they have a choice between an index-based hedge or a customised hedge. The metrics certainly need to move on from just looking at variance. As people have remarked, that has been used as a very simple way of assessing the differences and benefits of hedging.

If you are on the board of directors and, putting regulations to one side, you will look at things like economic capital. In particular, you will want to know how much less economic capital you need as a result of using a population-based hedge. You will have many other questions, for example, looking at economic capital over multiple time periods, applying a cost of capital method, or some

488

alternative, and quantifying how much of a benefit there is to the shareholders as a result of that. Obviously, the sorts of models that have been talked about are the way forward in that regard. Economic capital in its more general sense as well as the regulator's own measures of risk are where the focus should be shifting in the immediate future.

Prof Blake: The issue that I would like to raise is that of pricing transparency. Recall the early days after financial derivatives like currency options and interest rate swaps first started. We first saw academics working on theoretical models. Then we started getting transactions. There was close collaboration between academics and practitioners to see whether the prices of those transactions corresponded to the theoretical models. The markets took off that way and we quickly moved from mark-to-model pricing to mark-to-market pricing.

The situation is completely different in the new longevity derivatives market. There have now been dozens of longevity swaps traded. But, as academics, we know very little about either the contract terms or prices. Therefore, the important job of academics of analysing the swaps and assessing pricing efficiency is not currently happening. In order to help support the market, there needs to be much greater cooperation with academics and much more transparency over contract terms and pricing. I am, in particular, interested in mortality and longevity bonds. Some of these bonds have been issued, but there is very little public information on their pricing. Until that happens, this is going to slow down the development of the market, which we all want to develop as fast as possible.

The Chairman: Thank you very much. I will ask Mr Gaches to sum up from the panel's perspective and the authors' perspective.

Mr Gaches: I am just going to pick up on two or three of the themes on which perhaps we have not had much discussion, but they have certainly been alluded to by some of the comments.

Picking up on one of Professor Blake's comments on capital relief, and building on that, it is certainly the case that it is important for insurers to start to be able to understand what capital relief they could get it if they did reduce longevity risk through index-based transactions. It is also the case that in recent years there has been increasing focus on quantifying the various aspects of longevity risk. One of the aspects, which insurers have been asked to look at by the likes of the PRA is the question of basis risk. Clearly, we cannot speak for the regulator. But I would make a few comments. Insurers do need to use a robust approach to determine capital requirements. They are rightly subject to challenge if they do not do so. The better they are able to assess and articulate the risks that they hold, the greater confidence that they and indeed the regulator are going to be able to have in their assessment of capital requirements, and the better placed they are going to be to argue their case for what they think is appropriate capital. We very much hope that this research and the other research, which is going on in this area, advances the understanding of the modelling of longevity basis risk, and so contributes to insurers' abilities to make allowance for the benefits of de-risking with index-based solutions and also the allowance for basis risk in reserving.

The second area I wanted to comment on is picking up on one of the points made by Mr Ashmore. This alluded to what it will take for the future development of index-based longevity swap markets? How do potential participants go about assessing index-based swaps versus indemnity swaps?

It is true that from a UK pension scheme perspective, the market to date has been driven by bespoke transactions. That is not really surprising. We can draw parallels perhaps with financial de-risking

solutions which started out as largely bespoke. The pricing of the bespoke pension swaps that we have had to date it has been attractive so why would a scheme at the moment use an index-based solution, where the risk reduction is hard to assess, when a tailored approach to date has been as cheap if not cheaper? The key is that markets do evolve. Bespoke solutions do not meet all users' needs. They have not provided a mechanism for wholesale transfer of risk from non-pensioners. They do not provide a cost-effective short-term protection that some schemes do want. For example, schemes that are looking to buy out in 10 years' time may not want whole of life protection. The competitive pricing for bespoke deals may come under pressure as more schemes transact and the balance between supply and demand shifts. Structures that can draw in other risk-takers may be set to become more attractive. The challenge of assessing index-based solutions will diminish as the question of modelling basis risk is advanced. The £50 billion of bespoke longevity swaps undertaken to date has been a success. Ultimately, we need a solution that can manage two trillion pounds of liability. The standardisation of index-based structures will be needed at some point if the risk transfer market is ultimately to meet that challenge.

Finally, just a few thoughts on the general philosophy of using past data to assess basis risk. This is certainly a question which has been raised a number of times before. Clearly, it would be foolish to believe that the past allows us to predict the future. Nor can the past give us any certainty about the level of risk that the future holds. The thing is, the past is really the only guide that we have got. From my perspective, it would be even more foolish to try to form a view on the future without considering the past. What we have presented here is a model which builds on past data. We hope that practitioners find that this is a useful framework. Even more, we hope the practitioners use it in a way that I am sure regulators would want us to use it, with a view both to the strength and the weaknesses of the methods that we are proposing. We hope that this research will progress the industry's ability to assess basis risk, and support the continued success and growth of the ability of both pension schemes and insurers to manage longevity risk. Thank you.

The Chairman: Thank you to everyone for coming and for contributing to a very interesting discussion. I should remind you that the IFoA and the LLMA are considering the best way forward in relation to phase 2 of this project. All your comments will be taken into account for that.

May I ask you now to thank our speakers and our authors for the presentation.

References

IFoA/LLMA (2014a). Longevity basis risk. A methodology for assessing basis risk by Cass Business School and Hymans Robertson LLP, available at http://www.actuaries.org.uk/sites/all/files/IFoA%20LLMA%20Longevity%20Basis%20Risk%20Report_0.pdf

IFoA/LLMA (2014b). A methodology for assessing longevity basis risk. User guide, available at http://www.actuaries.org.uk/research-and-resources/documents/longevity-basis-risk-user-guide.