Institute and Faculty of Actuaries

Life & Longevity Markets Association

Longevity Basis Risk

A methodology for assessing basis risk

by Cass Business School and Hymans Robertson LLP

Research Report

8th December 2014

Cass Business School
CITY UNIVERSITY LONDON

HYMANS ROBERTSON

Longevity Basis Risk
A Methodology for Assessing Basis Risk

Research investigation and report by Cass Business School and Hymans Robertson LLP for the Institute and Faculty of Actuaries and the Life and Longevity Markets Association

Written by
Prof Steven Haberman FIA, Prof Vladimir Kaishev, Dr Pietro Millossovich, Andrés Villegas MACA

For and on behalf of Cass Business School

Steven Baxter FIA, Andrew Gaches FIA, Sveinn Gunnlaugsson GradStat, Mario Sison

For and on behalf of Hymans Robertson LLP

December 2014

Introduction

On behalf of the joint Longevity Basis Risk Working Group (LBRWG) established by the Life and Longevity Markets Association (LLMA) and the Institute and Faculty of Actuaries (IFoA), I am delighted to introduce the results of this research project.

This technical report details the methodology developed on behalf of the LBRWG to assess longevity basis risk. A user-guide which provides a high level summary of this report has also been produced. Together these documents form the key outputs of the first phase of a longevity basis risk project commissioned and funded by the IFoA and the LLMA, and undertaken on our behalf by Cass Business School and Hymans Robertson LLP.

The importance of longevity basis risk

Longevity basis risk arises because different populations, or subpopulations, will inevitably experience different longevity outcomes. This is a significant issue for those wishing to hedge longevity risk using a published mortality index – whether they be pension schemes, insurers, reinsurers or banks. To put it simply, actual longevity outcomes, and therefore cashflows, of the hedged portfolio will differ from those under the hedging instrument.

In addition, longevity basis risk can also present a wider issue for insurers using, in their reserving models, external data, such as population data, rather than their own policy data. The need to quantify and reserve for any potential basis risk is receiving increasing focus, particularly under Solvency II.

Demographic aspects of longevity basis risk

There are several aspects of longevity basis risk. This research focuses on the impact of demographic and socio-economic differences between the portfolio and the index population, which can lead to different initial rates and trends in mortality. To date, there has been no well-established methodology for assessing these demographic aspects of longevity basis risk.

Historical differences demonstrate the need to assess basis risk

A review of existing literature and analysis of pension scheme data have provided evidence that historic mortality improvement rates have varied by socio-economic class and deprivation. These variations have been significant and sometimes as large as the variation by gender. This demonstrates the significance of demographic basis risk and confirms the need to model longevity basis risk.

The need for a two-population model

To be able to assess demographic basis risk, the required model needs to able to capture the mortality trends in both the reference population backing the hedging instrument and in the population of the portfolio being hedged. Given this model, the assessment of other aspects of basis risk, such as sampling risk and structuring risk, becomes (in theory, at least) more straightforward.

Delivering a framework to assess longevity basis risk

I am delighted that the research has delivered a framework for assessing longevity basis risk. This recognises the fact that different users, with different portfolios, will have different constraints on the models they can use in practice. The research has identified specific models and techniques for different situations, which we believe will provide a good starting point for assessing basis risk.

We are delighted to be able to present this research and hope it will prove of value to practitioners and enable an important step change in the ability to assess longevity basis risk.

Pretty Sagoo

Chair of the LLMA and IFoA Joint Longevity Basis Risk Working Group

Scope, reliances and limitations

This report has been produced by Hymans Robertson LLP and Cass Business School for the Longevity Basis Risk Working Group (LBRWG) of the Institute & Faculty of Actuaries (IFoA) and the Life and Longevity Markets Association (LLMA).

The scope of this phase of work is limited to producing a proposed methodology for assessing (demographic) basis risk. For example identification and development of appropriate metrics for assessing basis risk, quantification of potential capital savings and presentation of basis risk results to regulatory authorities are excluded from this initial phase and (potentially) form part of a secondary phase of this project.

This report is addressed to the LBRWG. It may be shared with members of the IFoA and LLMA and other relevant third parties. This report does not constitute advice and should not be considered a substitute for specific advice in relation to individual circumstances. While care has been taken to ensure that it is accurate, up to date and useful, neither Hymans Robertson LLP, Cass Business School, the IFoA nor the LLMA accept liability for actions taken by third parties as a consequence of the information contained in this report.

Executive Summary

This paper summarises the work to date of Cass Business School and Hymans Robertson LLP in relation to assessing longevity basis risk. This work was commissioned by the Longevity Basis Risk Working Group (LBRWG) and funded by the Life and Longevity Markets Association (LLMA) and Institute and Faculty of Actuaries (IFoA). The LBRWG was formed by the LLMA and IFoA in December 2011 with a remit to investigate how to provide a market-friendly means of analysing longevity basis risk.

The key outputs of this work are:

- for modelling books which are 'self-credible' (i.e. a large number of lives & sufficient back history) a shortlist of 'best of breed' 2 –population models (specifically the **M7-M5 model**, or in some situations the **CAE+Cohorts model**);
- for modelling the majority of books which are not self-credible, an alternative, easy to apply "**characterisation approach**";
- a clear **decision tree framework** to aid the selection of an appropriate methodology for assessing basis risk from those mentioned above;
- a **clear recognition of the importance of choice of time series** underpinning any 2- (or multi-) population model

These outputs are backed up by an extensive body of research, including:

- a review of **how trends have varied between different (sub) populations** in the past, covering both the highlights of existing literature and additional research based on the Club Vita dataset of UK occupational pension schemes;
- a review, classification and general formulation of two-population models that could be considered for modelling longevity basis risk;
- a **thorough and systematic assessment of candidate two-population mortality models** to identify those which provide the most suitable balance between flexibility, simplicity, parsimony, goodness-of-fit to data and robustness;
- **case studies, review of key challenges and consideration of practical issues** in relation to both the M7-M5 model and the characterisation approach.

Introduction to longevity basis risk

When insurance companies and pension schemes consider managing their longevity risk one of the available options is to use a hedging instrument based upon published mortality indices. However this has a risk that the actual longevity outcomes (and so cashflows) of the hedged portfolio may differ from those under the hedging instrument.

This may happen due to structuring risk (i.e. the hedging instrument having a different payoff structure to the hedged portfolio), sampling risk (arising from the different random outcomes of the individual lives within the portfolio and the index population) or demographic risk (with demographic and socio-economic differences in the composition of the portfolio and the index population leading to different initial rates of mortality and trends therein). This project focuses specifically on the question of demographic risk which has no well-established assessment methodology.

Historical differences in mortality improvement rates demonstrate the need to assess basis risk

Our review of existing literature demonstrates clearly that mortality improvement rates have historically varied by socio-economic class and deprivation. These variations have been significant – indeed they have been as large as the variation seen by gender. This conclusion is confirmed by analysis of the trends in the Club Vita dataset of occupational pension schemes.

The size of these variations demonstrates the significance of demographic basis risk and confirms the need to model longevity basis risk.

The need for a two-population model

In order to be able to assess basis risk, we need a model that is able to capture the mortality trends in the reference population backing the hedging instrument and in the book population, the longevity risk of which is to be hedged. This modelling is needed in order to generate a distribution of future scenarios to evaluate the possibly different evolution of mortality in the two populations. Given this model, the assessment of sampling risk and structuring risk becomes (in theory, at least) straightforward.

Directly modelling basis risk

If a book is 'self-credible' (i.e. has large number of lives and sufficient back history) it is possible to parameterise a two-population model directly from mortality experience data relating to i) the population underlying the index and ii) the book population.

Our systematic assessment of candidate two-population mortality models identified two particular 'best of breed' two–population models (specifically the M7-M5 model, or in some situations the CAE+Cohorts model).

Parametric form for shape of mortality with age ('M7-M5')

The M7-M5 model is a two-population extension of the Cairns-Blake-Dowd (CBD) model of mortality introduced in Cairns, Blake, & Dowd (2006).

Readers may be familiar with the Cairns-Blake-Dowd family whereby the logit of mortality (as measured by q_x) takes up to a quadratic form with age. Within this family of models we find incorporating both a quadratic term (to capture shape sensitivities at older ages) and a cohort term leads to the best performance for the reference population, and results in the model often referred to as 'M7' as per Cairns et al (2009):

Thus the (M7) model for the reference population R takes the form:

$$\text{logit}\, q_{xt}^{R} = \kappa_t^{(1,\text{R})} + (x - \bar{x})\kappa_t^{(2,\text{R})} + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^{(3,\text{R})} + \gamma_{t-x}^{R}$$

The difference between book population B and reference population R takes the form of a simplified Cairns-Blake-Dowd model, with linear age sensitivity and no cohort effect, often referred to as 'M5'.

Hence the (M5) model for the difference between book population B and reference population R takes the form:

$$\text{logit}\, \text{q}_{\text{xt}}^{\text{B}} - \text{logit}\, \text{q}_{\text{xt}}^{\text{R}} = \kappa_{\text{t}}^{(1,\text{B})} + (\text{x} - \bar{\text{x}})\kappa_{\text{t}}^{(2,\text{B})}$$

Non-parametric form for shape of mortality with age ('CAE+cohorts')

If we instead allow the shape of mortality to have a non-parametric relationship with age we obtain the extended Lee-Carter family of models. Within these models we find a Lee-Carter model with the addition of a cohort term performs best for the reference population

Thus the (Lee-Carter with cohort term) model for reference population R takes the form:

$$\text{logit}\, q_{xt}^{R} = \alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R$$

The difference between book population B and reference population R takes the form

$$\text{logit}\, q_{xt}^{B} - \text{logit}\, q_{xt}^{R} = \alpha_x^B + \beta_x^R \kappa_t^B$$

This model is referred to as the common age effect as the book population has a Lee-Carter form with the same sensitivity by age to time based improvements as the reference population.

Indirectly modelling basis risk – Characterisation approach

If the book is not 'self-credible', (i.e. it does not have a sufficiently large number of lives or lacks a sufficient back history) then it is not possible to robustly parameterise the book element of a two-population model directly from mortality experience data. In this situation an alternative approach is required.

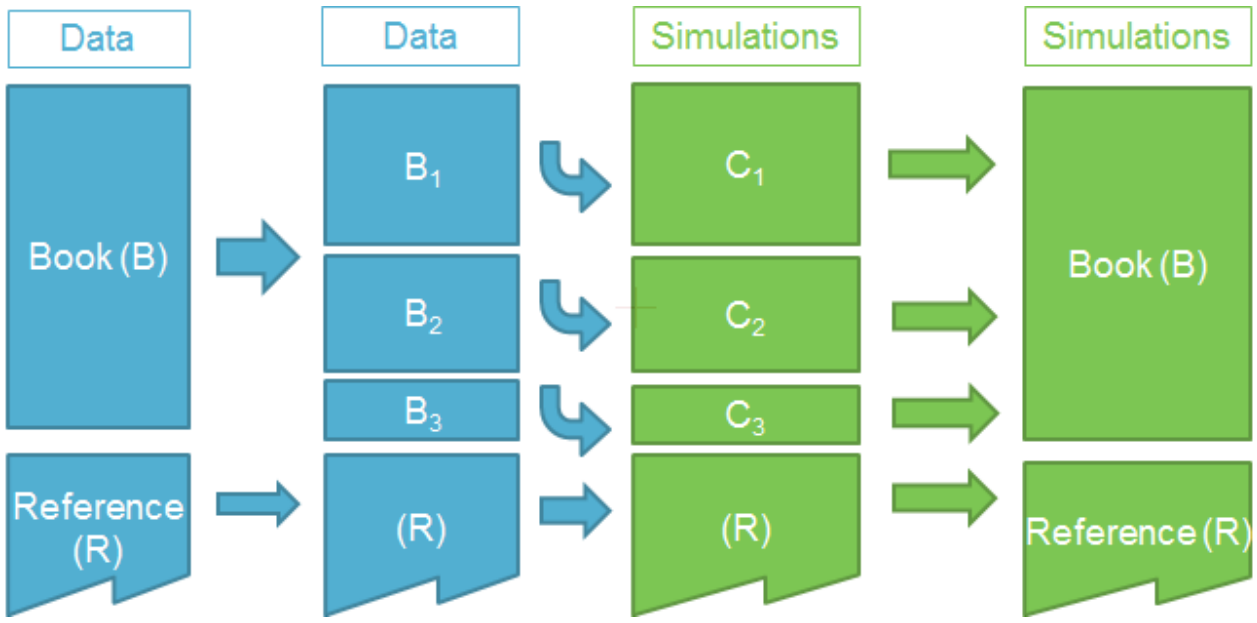
Indeed, even where the book is sufficiently large with long enough experience history to use direct modelling, an alternative indirect approach may still be useful; either as a means of a pragmatic initial assessment of the quantum of basis risk, or as an alternative approach as part of considering model risk.

The alternative we propose, which we describe as a "characterisation approach" enables an assessment of basis risk based on the *characteristics* of the book in question; leveraging an alternative larger dataset to provide the required volumes and back history of data.

Instead of using the experience data of the book itself, the basic principle of the characterisation approach is to map the book onto a small number of characterising groups which:

- capture the majority of the source of demographic risk
- can be projected using an alternative data source with a more reliable and longer back-history of mortality experience

Schematically, this approach can be thought of as:



In the schematic above, the book population B is subdivided into three distinct subgroups B_1, B_2 and B_3, according to some characterising criteria. Both B and the subpopulations B_1, B_2 and B_3 are too small for direct modelling. However, a larger characterising population C is available, and has previously been segmented (using the same characterising criteria) into subgroups C_1, C_2 and C_3. Importantly, C has been chosen such that C_1, C_2 and C_3 are sufficiently large for direct modelling (in conjunction with the reference population R).

It is now possible to simulate B indirectly, by first simulating C_1, C_2 and $C_3$and mapping those simulations across to B_1, B_2 and B_3.

Choosing between approaches

A flow chart (see next page) has been developed to assist users (based on their requirements) in choosing between direct modelling and the characterisation approach. In addition, it provides assistance in choosing between M7-M5 and CAE+Cohorts models where direct modelling is preferred.

Having assessed direct modelling under M7-M5 and CAE+Cohorts models for book populations of different size and history length, a key requirement for direct modelling (reflected in the first question) is sufficient data; typically over 25,000 lives and in excess of 8 years history.

In addition direct modelling relies on the assumption that "past data is a good guide to the future". This may not always be the case, hence the second question relating to whether there have been any major changes in the socio-economic mix in the book over time.

There are a number of considerations which could be taken into account in choosing between M7-M5 and CAE+Cohorts (including user familiarity or preference), but a specific practical issue, relating to the typical need to allow for inter-age mortality correlations is covered by the third question.

Finally, in some cases there could be a strong belief in a book specific cohort effect (which would require an extension to the form of the model for book population); this is covered by question 4.

Case studies, key challenges and practicalities

Case studies are provided for both direct modelling and the characterisation approach. In addition we seek to identify (and suggest approaches to tackle) key challenge and practicalities in the application of these methods.

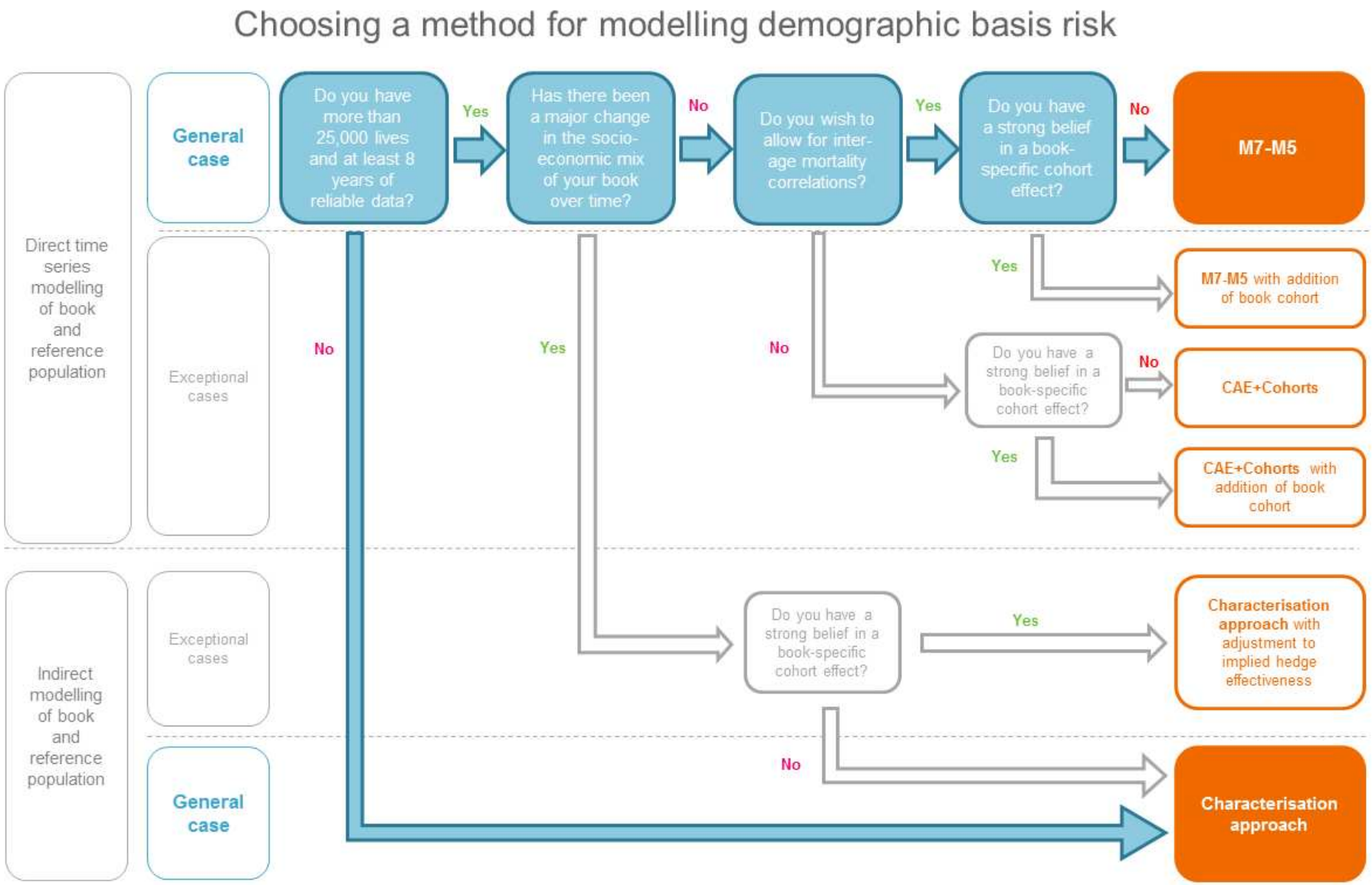
Sensitivity to choice of model and choice of time series

The alternative methods illustrated (M7-M5 and CAE+Cohorts for direct modelling, and the characterisation approach) provide for the most part similar conclusions on the amounts of basis risk. But differences do exist, illustrating the issue of model risk.

We focus on a number of well-established choices of time series in the models; the alternatives used demonstrate the model risk associated with choice of time series; a comprehensive exploration of alternatives and their impact is outside the scope of this research project. Nonetheless it is appropriate to flag the risk associated with choice of time series and highlight the benefit that further assessment, development and guidance on the choice of time series would bring to practitioners.

Nature of this paper

Please note this paper is designed to provide sufficient detail for a knowledgeable user to understand the methods we are proposing and the reasoning why we have chosen those methods. As such it is necessarily technical in places. Readers seeking a high level overview of the methodology and the key considerations in applying this method are directed to our accompanying user guide.


Choosing a method for modelling demographic basis risk
Direct time series modelling of book and reference population
General case
Exceptional cases
Indirect modelling of book and reference population
Exceptional cases
General case
Do you have more than 25,000 lives and at least 8 years of reliable data?
Yes
Has there been a major change in the socio-economic mix of your book over time?
No
Do you wish to allow for inter-age mortality correlations?
Yes
Do you have a strong belief in a book-specific cohort effect?
No
M7-M5
No
Yes
No
Yes
No
Yes
M7-M5 with addition of book cohort
Do you have a strong belief in a book-specific cohort effect?
No
CAE+Cohorts
Yes
CAE+Cohorts with addition of book cohort
Do you have a strong belief in a book-specific cohort effect?
Yes
Characterisation approach with adjustment to implied hedge effectiveness
No
Characterisation approach

Contents

Longevity Basis Risk PAGE

1 Introduction

1.1 Longevity risk transfer market

Recent years have seen a huge growth in longevity risk transfer, both in the insurer to reinsurer market, and from pension schemes to the insurance market[1]. An effective, growing market with sufficient capacity to meet demand would be to the benefit of all participants, whether to enable business to be done, or to manage risk.

To date most transactions have been "bespoke" deals, with the payouts linked directly to the actual experience or lifespans of the individuals being covered. But index-based solutions – where the payouts are linked to a longevity index or metric based on an external reference population – are possible. They have the potential to provide important benefits: lower costs, faster execution, potential for liquidity, and greater transparency.

1.2 Enabling the development of index based solutions

Many steps have been taken to enable index-based solutions to develop. Publication of indices by the LLMA, Deutsche Börse and others; continued innovation of possible structures such as the Longevity Experience Options introduced by Deutsche Bank; and papers on standard derivative structures such as q- and S-forwards[2].

But one key issue remains – that of "longevity basis risk". How good a match will there be between a portfolio's experience, and that reflected by an external, published, longevity index? How much protection can index-based solutions provide?

1.3 The question of longevity basis risk

In its simplest form an index based longevity swap involves a payment to the pension scheme or insurer that is based on the longevity experience of a reference index. An index–based swap provides a means to obtain (partial) protection from longevity risk both for pensioners but also deferred pensioners who are generally not covered by the "bespoke" transactions. In the case of life insurers they offer a potentially flexible way to manage exposure to longevity risk, or to facilitate a more capitally optimal balance between longevity and mortality risk. However index-based swaps do not provide a perfect risk reduction. The index based payments will not exactly match the actual annuity payments being made by the insurer or pension scheme.

Understanding the residual longevity risk and "how good" the risk reduction is, is key. The kinds of practical questions asked about index-based swaps include:

- What is the risk that index payments will fall short of annuity payments?
- How can we determine the "hedge effectiveness"?
- How can we do a cost-benefit analysis of an index-based hedge?
- How do we determine an appropriate capital reduction for an index-based hedge?

To answer these questions, we need a practical and realistic way of modelling and quantifying basis risk.

[1] For example in the year to 30 June 2014 £39bn of longevity risk was transferred from pension schemes to insurers and reinsurers via buy-ins, buy-outs and longevity swaps. Of this £27bn related to longevity only transactions (longevity swaps), close to double the volume written in the preceding 4 years. (Hymans Robertson (2014))

[2] See for example http://www.llma.org/publications.html

1.4 Sources of basis risk

There are three primary sources of basis risk[3]:

1 **Structuring risk** due to the pay-off of the hedging instruments being different to that of the portfolio (for example the hedging instrument making annual payments whereas the portfolio pays annuities or pensions monthly, or the hedge may be of shorter duration than the liabilities).

2 **Sampling risk** arising from the random outcomes of the individual lives within the portfolio and the index population meaning the actual mortality experienced by the two populations will not be the same, other than by chance.

3 **Demographic risk** owing to demographic and socio-economic differences in the composition of the actual portfolio being hedged and the index population referenced in the hedge, leading to different underlying mortality rates now - and in the future.

Well-established methods for modelling the first two of these exist. Structuring risk can be assessed by simulating the cashflows under the portfolio and the payoffs under the instrument, whilst sampling risk can be modelled by simulating the outcomes for the respective populations.

1.5 Demographic risk

In contrast there is no well-established methodology for assessing demographic risk. Yet it is this risk which worries (re)insurers and pension schemes when they consider entering index-based longevity transactions. The absence of a method for quantifying such risk makes it very difficult to assess whether such a transaction looks good value for money, or what impact the transaction would have on the insurer's or pension scheme's overall risk profile and hence capital / funding requirements. Our research is focused on this demographic aspect of longevity basis risk.

When considering a transaction we will know certain things about the portfolio: size, affluence, locations, maybe historical mortality experience. How then do we model the portfolio (and the reference population) in order to assess basis risk?

The key question that we explore is:

"What is an appropriate model for the mortality rates over time in the two populations?"

1.6 Scope of this research

This paper provides a detailed summary of the key elements of the work undertaken by Hymans Robertson and Cass Business School for the first phase of a research project commissioned by a joint Longevity Basis Risk Working Group (LBRWG) of the Institute & Faculty of Actuaries (IFoA) and of the Life and Longevity Markets Association (LLMA) aimed at answering the above question.

Fuller details of the LBRWG's call for proposals are provided as Appendix F. In summary that call split the project into two phases, with commissioning of Phase 2 dependent upon completion of Phase 1.

Phase 1

Provision of:

- Review of evidence of different mortality improvement rates among different subgroups to inform projection methodology (sections 2 & 3 of this report)

[3] As described by Mosher & Sagoo (2011)

- Critical review of existing models for relationship between a specific book (portfolio) and reference population mortality (sections 4, 5 & 6 of this report)
- Detailed specification of a proposed methodology (sections 7 & 10)
- Analysis of the limitations of the methodology and description of alternatives (sections 9 & 12; building on sections 6 & 7)
- Clear specification of work to be completed and anticipated outputs of Phase 2 (not covered in this report)

Phase 2

- Identification of basis risk metrics covered by the proposed model and demonstration of how outputs of methodology can be used for these metrics
- Application of the model on practical, realistic, illustrative examples based on the data reasonably available to potential users (sections 7 & 10, and Appendix D, of this report provide some initial case studies)
- Demonstration of how outputs from the methodology can be presented as robust quantification of basis risk to third parties such as regulators

1.7 Structure of this paper

This report covers the analyses carried out by the team in response to Phase 1 and includes a proposed methodology. We start by considering what history tells us about longevity basis risk. Section 2 summarises how trends have varied between different (sub) populations in the past, and section 3 provides a high level review of the drivers of those trends. This context informs the choice of models and the way in which users ultimately apply and interpret any results.

Sections 4 and 5 set out the modelling problem more formally and provide an overview of the models available to tackle the question at hand.

In section 6, we summarise the steps taken to narrow down the wide range of models to those likely to be most useful to practitioners. Section 7 explores in more detail the two main contenders identified, including their strengths and weaknesses, and proposes a decision tree suggesting which modelling approach may be a good starting point in different situations. An illustrative case study of the approach where the user can rely on the portfolio experience data ('direct modelling') is provided in section 8.

In many cases the portfolio experience data will be insufficient to calibrate models directly and so alternative techniques are required. Before we move on to these alternative techniques, section 9 reviews some of the key challenges and addresses several practical questions on the use of the direct modelling techniques described so far. This section in particular moves the debate on from choice of model (e.g. Cairns-Blake-Dowd) to highlighting the need for users to consider the type of time series driving these models.

In section 10, the focus moves away from modelling the reference and book populations directly, and considers a more indirect approach whereby the book population is "characterised" into buckets for which alternative datasets can provide a measure of demographic basis risk. As such it extends the scope of Phase 1 in order to create a methodology with wide practical application. A case study of this "characterisation approach" is provided in section 11, with section 12 considering the practical issues associated with a characterisation approach.

Finally, it is anticipated that a subsequent phase of the project (and reports) will look at such issues as appropriate metrics for quantifying the basis risk and further back testing of the realised vs predicted hedge effectiveness for a range of portfolios. As such further work may lead to some refinements to the conclusions drawn here, however we believe this work provides an appropriate starting point for those seeking to assess basis risk.

2 Observed differences in mortality improvements

Differences in baseline mortality are generally well understood by practitioners who are well-versed in allowing for these differences. There are many sources of evidence – both from the UK population as a whole, and from the experience of pensioners and annuitants – that show very significant differences (up to approximately 10 years) in lifespans for different types of individuals (ONS (2014); Madrigal et al. (2012)).

Differences in historically observed improvement rates between populations are also well known by the life and pensions industry but less commonly modelled prospectively. In this section, we review existing published evidence on differences in observed improvements (section 2.1) and provide additional, new, results specific to the experience of pension scheme annuitants using the Club Vita dataset (section 2.2; further details on dataset in Appendix A).

2.1 Existing research

2.1.1 Improvement differentials by gender

Figure 2.1 shows the average annual improvement rate over a 30 year period, for men and women from the England & Wales population at various ages.

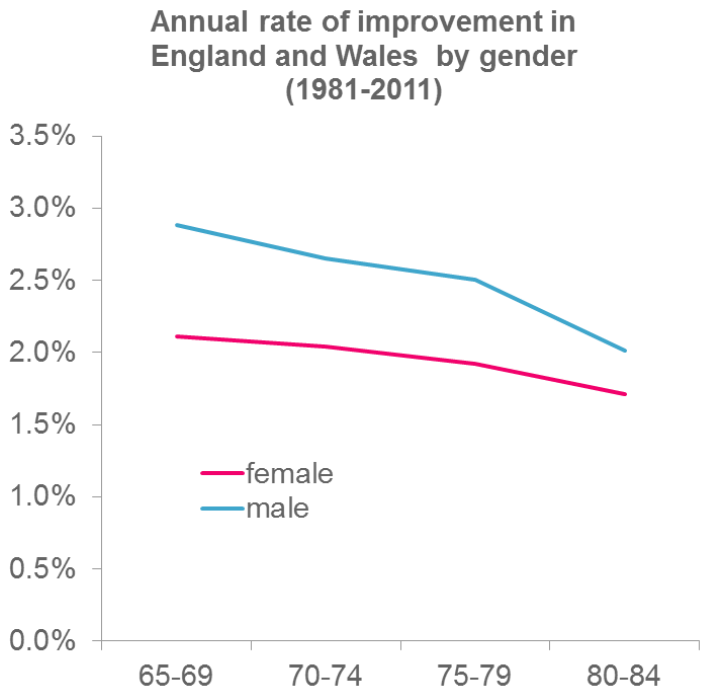


Figure 2.1: Annual rate of improvements in England and Wales by gender (1981-2011) based on HMD data.

We can see clear and consistent differences; with men having experienced annual improvements around 0.5%-0.75% higher than women between 1981 and 2011. These faster improvements for men underpin the well-known closing of the life expectancy gap between men and women. However, existing longevity indices provide separate values for men and women; as such, these are differences in improvements which, if replicated in the future, can be hedged and so are unlikely to be a source of longevity basis risk in practice.

2.1.2 Improvement differentials by deprivation

In contrast there are currently limited options to access longevity indices which differentiate by socio-economic status. However, the differences in improvements seen historically for different socio-economic groups are comparable to those seen between men and women. We can see this by, for example, looking at improvements by deprivation.

There are various options for measuring deprivation, including Townsend's and Carstairs' index and the Index of Multiple Deprivation (IMD). We have focused on IMD in our analysis and used the 2007 version.

IMD 2007 combines indicators across seven deprivation domains (e.g. income, employment, health, education, crime rates, etc.) into a single deprivation score. These scores are available for a range of geographical regions. In our analysis we have focussed on the scores for Lower Layer Super Output Areas (LSOAs), each of which have an average of 1,500 residents and around 650 households[4].

The LSOAs are ranked by their IMD score and grouped into quintiles where Q1 represents the least deprived areas and Q5 the most deprived areas[5]. Figure 2.2 shows the observed annual improvement rates within England for each quintile, over a similar period as shown for men and women in section 2.1.1.

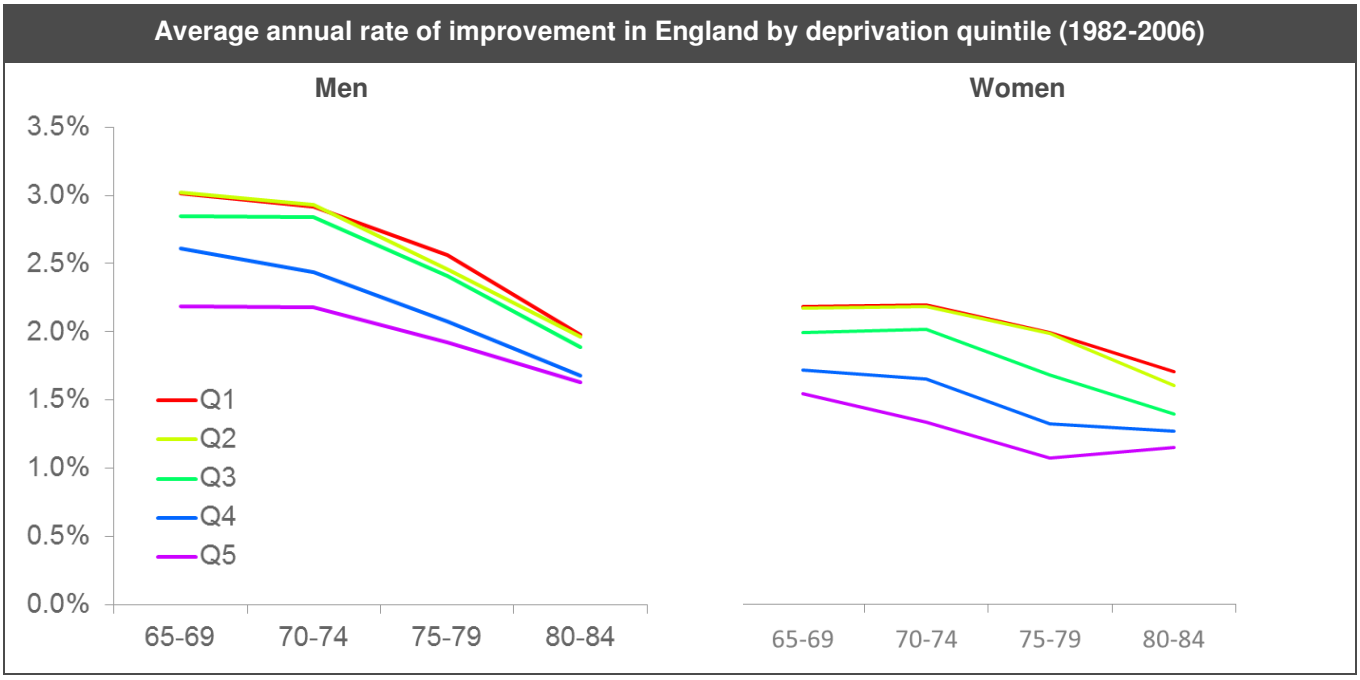


Figure 2.2: Annualised improvements in mortality, England by deprivation quintile. Based on Table 1 and 2 in Lu et al. (2013)

We can see how for:

- **Men:** The least deprived areas (Q1 red line) have experienced average annual improvements of around 0.5-0.75% higher than the most deprived areas (Q5 in purple).
- **Women:** The improvements are generally lower (consistent with Figure 2.1) but the pattern and spread is very similar as for men.

Notice how the differences in improvements between the least and most deprived areas are of a very similar magnitude to the differences between men and women shown in Figure 2.1. In the context of longevity risk, this

[4] http://neighbourhood.statistics.gov.uk/HTMLDocs/nessgeography/superoutputareasexplained/output-areas-explained.htm

[5] For further information, see (Indices of Deprivation 2007 for Super Output areas, 2007).

highlights the potential for index-based solutions to provide a less than perfect hedge, and hence the need to be able to quantify demographic basis risk.

2.1.3 Improvement differentials by condensed NS-SEC (Socio economic class)

Figure 2.3 shows the improvement rates in the England & Wales male population by an alternative measure of socio-economics; the condensed NS-SEC[6,7] (National Statistics Socio-Economic Classification).

We can see how the managerial & professional groups experienced the highest annual rate of improvements at most older ages and the routine & manual group the lowest. The data here is more volatile as it is based on the 1% sample in the ONS longitudinal study; but again there is a clear difference in past annual improvement rates (1% on average) and hence in trends over that period.

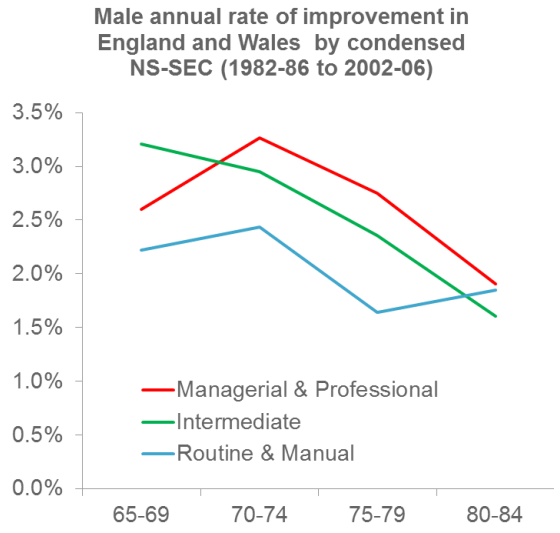


Figure 2.3: Annualised mortality improvements by NS-SEC.
Source: ONS Longitudinal Study

2.1.4 Mortality differentials by income

Various studies have explored how improvements in mortality rates differ by income.

For example evidence of the potential for improvements to vary by affluence is provided by Adams (2012) which analysed mortality differences by pension income in Canada between 1993 and 2007. Pension income was split into five (non-distinct) classes based upon the maximum level:

- **Class 1:** <35% Maximum pension *(omitted from charts below in original paper)*
- **Class 2:** 35% - 94% Maximum pension
- **Class 3:** 95% - 100% Maximum pension
- **Class 4:** 35% - 100% Maximum pension
- **Class 5:** All income

The figures below extracted from that paper, show the (fitted) annualised mortality improvement for each of income classes 2, 3, 4, and 5 by different age for men and women. Looking at the chart for men, it is clear how class 3 (which represents the most affluent group) shows the highest annual improvements, in particular when focusing on ages 60-80. We also see how income class 2 (which represents those on the lower end of the affluence spectrum) has shown relatively lower improvement rates over most of the post-retirement age range.

The results for women (figure 2.5) are a lot more volatile but similar trends appear as in the case for men where income class 3 (highest pensions) seems to have the highest annual improvement rates between ages 60 and 80.

[6] See The National Statistics Socio-economic classification (2010).

[7] Note that the National Statistics moved to a revised new measure in year 2000, SOC 2000 from SOC 90 (See Rose & Pevalin (2005)), which limits our ability to look beyond 2006 in this analysis.

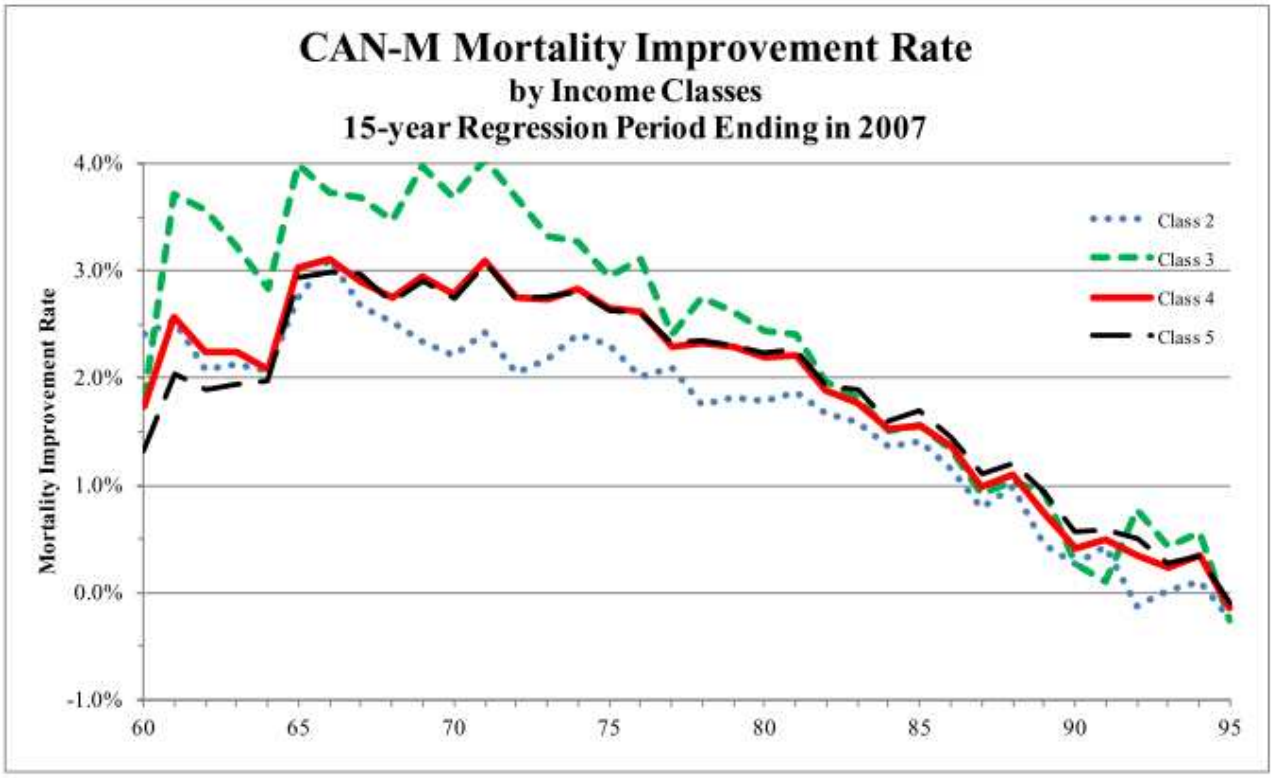


Figure 2.4: Annualised mortality improvements by income class for Canadian men, at various ages, between 1993 and 2007; extracted from Adams (2012)

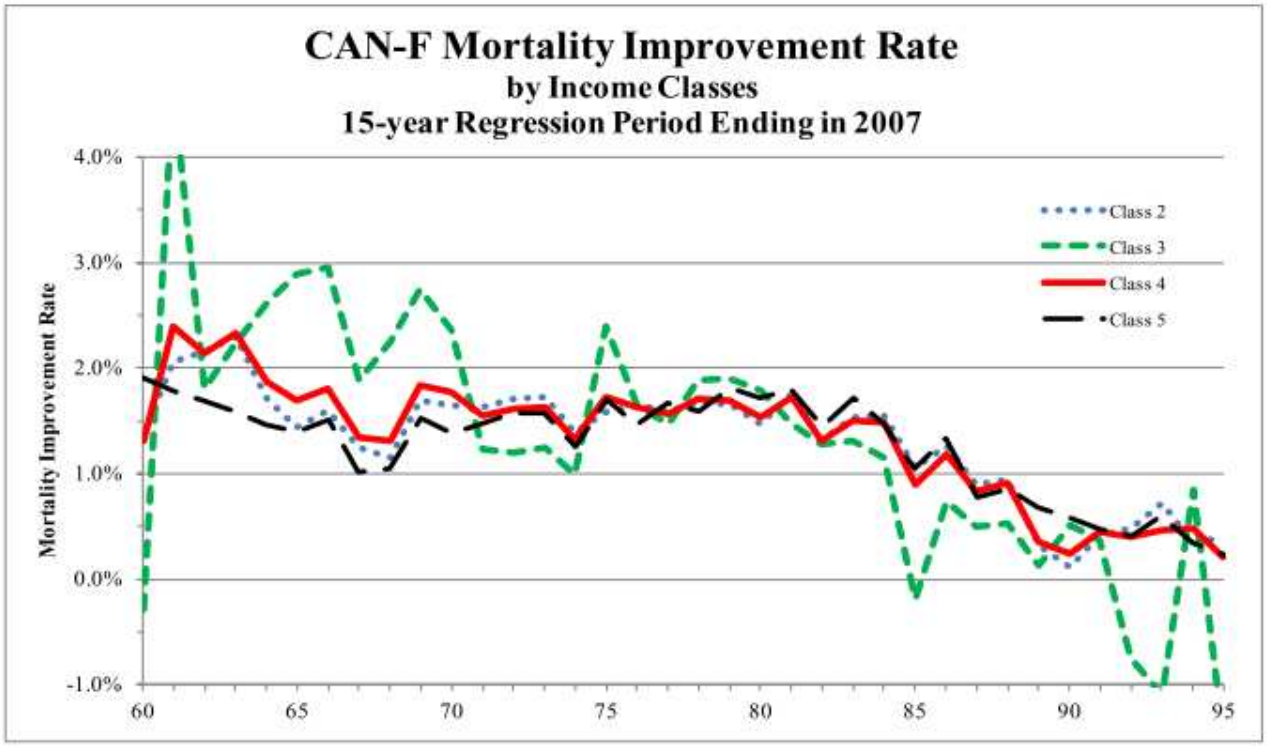


Figure 2.5: Annualised mortality improvements by income class for Canadian women, at various ages, between 1993 and 2007, extracted from Adams (2012)

2.2 Separating improvements by socio-economic factors

The previous section has highlighted some clear differences in mortality improvements when segmenting national data by gender, income, deprivation quintiles or socio-economic classes. Each of the analyses presented shows the improvements across a single one of these dimensions (within gender). However, these analyses cannot be combined owing to the interrelated nature of deprivation, socio-economics and affluence. In the context of modelling demographic risk it would be more insightful if we could understand which of these factors are most predictive of historical trends, as this might indicate the factors which are most relevant to future demographic risk. To do this, we need to turn to a more granular dataset.

One such dataset is Club Vita, which holds detailed information (postcode, affluence, occupation, etc.) for living and deceased members of UK occupational pension schemes. (See Appendix A for further information.)

2.2.1 A model to identify the key predictors of historical improvements

To identify the main characteristics that explain differentials in mortality improvements, we have carried out a multivariate analysis of observed historical improvements within the Club Vita dataset, separately for men and women. The aim of this analysis is to identify which factors are closely linked to strong differences in historical improvements rather than optimising the best possible model. As such a simplified model was constructed using the framework of GLMs (generalised linear models), with a logit link function under a binomial setting.

Specifically, we carried out a two-step process to fit to the observed improvement data[8]:

1 Fit a baseline model as a linear function of the key mortality predictors identified by Club Vita; age, retirement health, pension amount, postcode based lifestyle[9], and IMD deprivation quintile i.e.

$$logit\left(q_{xtijkl}\right) = \alpha + \beta_0 x + a_i^{(0)} + a_i^{(1)} x + b_j^{(0)} + b_j^{(1)} x + c_k^{(0)} + c_k^{(1)} x + d_k^{(0)} + d_k^{(1)} x$$

Where:

- q_{xtijkl} is the one year mortality rate for a person age x at time t belonging to healh status group i, pension band j, lifestyle k, and IMD quintile l
- $logit\left(q_{xtijkl}\right) = log\left(\frac{q_{xtijkl}}{1-q_{xtijkl}}\right)$
- $\alpha + \beta_0 x$ describes the average level of logit mortality with respect to age x as linear with age
- The terms $a_i^{(0)}, b_j^{(0)}, c_k^{(0)},$ and $d_k^{(0)}$ determine the relative adjustment[10] to the intercept for someone belonging to retirement health status group i, pension band j, lifestyle k, and IMD quintile l, respectively
- The terms $a_i^{(1)}, b_j^{(1)}, c_k^{(1)},$ and $d_k^{(1)}$ determine the relative adjustment to the linear relationship for someone belonging to retirement health status group i, pension band j, lifestyle k, and IMD quintile l, respectively

This provides a proxy to the general industry approach of using a granular model to capture the baseline for a portfolio, and was fitted to data spanning 1993-2011. By incorporating pension amount within this model we also (broadly) capture the impact of amounts vs lives weighted mortality.

[8] The data ranged from 1993 to 2011, focusing on pensioners aged from 65 to 94 living in England only (to enable use of postcode based deprivation scores). See Appendix A for more details on data used

[9] Using Club Vita's proprietary postcode based lifestyle rating factors – see Appendix A

[10] Relative to the reference level of each predictor i.e. where $a_i^{(0)}, b_j^{(0)}, c_k^{(0)}, d_l^{(0)}$ and $a_i^{(1)}, b_j^{(1)}, c_k^{(1)}, d_l^{(1)}$ are fixed at 0 for a particular reference group

2 The resulting baseline model was then extended by adding mortality improvements to the models, conditional on the already fitted baseline parameters i.e.

$$logit(q_{xtijkl}) = \text{Baseline} + \tau_x t + t(b_j^{(2)} + c_k^{(2)} + d_l^{(2)})$$

Where:

- Baseline is derived from step 1
- t is our time index (with $t = 0$ corresponding to 1993)
- τ_x is the average annual improvement observed at age x
- The terms $b_j^{(2)}, c_k^{(2)},$ and $d_k^{(2)}$ determine the relative adjustment to the annual improvement for someone belonging to pension band j, lifestyle k, and IMD quintile l, respectively

As the aim here is to investigate the relative importance of the variables with respect to mortality improvements in a simple way, rather than coming up with the 'perfect' model for historic improvements, we have omitted the cohort effect. This has the benefit of substantially simplifying the construct of the model (as removes identifiability issues), the parameter estimation and the interpretation of the results.

Step 2 was repeated varying which predictors were included in the improvements component. When considering available covariates to be included in this component we focused on those predictors which are generally available to the industry. By doing so we ensure that the results have a practical application. For example, pension amount was chosen as the affluence covariate (instead of salary amount) due to its wider availability. Within this analysis we retained two postcode based metrics – one based upon publicly available deprivation scores, and the other using Club Vita's 'lifestyle' groupings based upon ACORN classification. Whilst the later of these is not publicly available, it is included in this analysis as a proxy to the proprietary postcode based lifestyle proxies used by many practitioners.

Table 2.1 shows the results of this analysis, identifying the most significant rating factors, in relation to past improvements, after having penalised for extra complexity of introducing additional parameters to the model[11].

[11] When penalising for extra parameter complexity in the model the AIC (Akaike Information Criteria) was used which is defined as $2(n - \log(L))$ where n is the number of parameters in the model and L is the likelihood under the model of the observed values (here improvements)

Variables included in fitting improvements	Rank Men	Δ AIC[12] Men	Rank Women	Δ AIC[13] Women	Comments
Deprivation (IMD via postcode) + Pension amount	1	0	2	1[14]	Best fit for men and (essentially) for women
Deprivation (IMD via postcode)	2	4	1	0	Narrowly best fit for women
Lifestyle (via postcode) + Pension amount	4	21	3	10	
Lifestyle (via postcode)	6	28	4	11	
Pension amount	3	16	5	13	
"No specific improvements predictor"	5	24	6	14	The worst fit is given by a model with no allowance for socio-economic factors

Table 2.1: Results of GLM analysis of fitting to historical mortality improvements within the Club Vita dataset

Reassuringly, the results for men and women give a very consistent message, that a model allowing for both the IMD deprivation index and pension amount[15] provides the best balance between fit to historical improvements and simplicity.

2.2.2 The relationship between the key predictors and improvements

We can look at the best fitting model above in more detail. Specifically it provides three *additive* terms which we can look at in turn:

- a general level of annual improvement at each age;
- an adjustment depending on deprivation; and
- an adjustment based on pension amount.

So, for example to calculate the annual rate of improvement applicable to men aged 70-74, living in an area within IMD quintile 2 and with a pension over £10,000pa we can add together the values shown in each of the charts in this section.

[12] Reference AIC for men (Postcode (IMD) + Pension amount) was **760,477.5**

[13] Reference AIC for women (Postcode (IMD)) was **353,059.4**

[14] Although a slightly higher AIC than for deprivation without postcode the difference is sufficiently small that the two models can essentially be considered equally good

[15] When looking at the above table, note that the pension amount has 3 levels, whilst the postcode metrics have 5 levels. All else being equal this would favour ranking postcode metrics higher up the table than pension based metrics. We initially considered using five levels for pension to ensure comparable numbers of levels, but regrouped into three groups on grounds of parsimony as there were little difference between the first three quintiles both in terms of improvements and mortality levels. It is reassuring therefore to see that adding pension to a postcode based model improves fit, confirming it is beneficial to include both pension and deprivation.

2.2.2.1 Age

Figure 2.6 shows how the fitted mortality improvement rates vary with increasing age, for men and women. In each case the solid line is the fitted value, and the dotted lines a 95% confidence interval based on the uncertainty in the fitted parameters.

We can see how:

- improvements decline with increasing age for both men and women.
- the improvement rates at the oldest ages have not completely converged to zero which indicates that we are still observing significant improvements in mortality at the oldest ages.
- mortality improvements have generally been lower for women than men, which is consistent with previous results from England and Wales data in Figure 2.1.

  (We observe more uncertainty in the average rates for women as shown by the wider confidence intervals due to smaller population size)

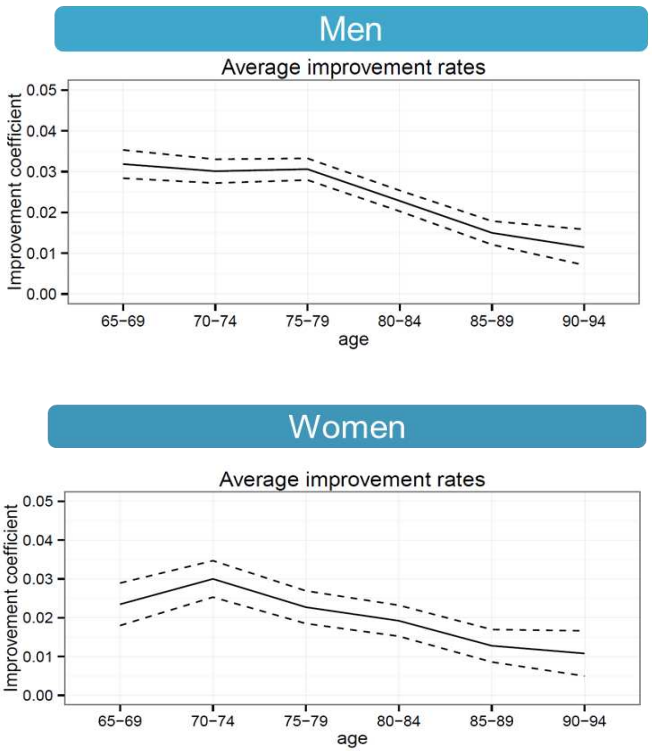


Figure 2.6: Age component of fitted improvements

2.2.2.2 Adjustment for deprivation (IMD 2007)

Figure 2.7 demonstrates the relationship between the IMD deprivation index and the historical mortality improvement rates, relative to the average level of deprivation, for men and women.

This suggests that:

- those living in the most deprived areas have experienced significantly faster improvements than those living in average deprivation areas (as indicated by the confidence interval for the improvements for Q5 excluding 0); whilst
- those living in less deprived areas have had similar improvement rates on average, especially for men.

Figure 2.7: Deprivation component of fitted annual improvements (relative to average level across all deprivations)

2.2.2.3 A deprivation paradox?

By comparing historical improvements by IMD for occupation pension members to what has been observed on the national level we observe very different trends. Figure 2.8 below compares the results for men from the multivariate analysis of Club Vita data with the univariate analysis of Lu et al. (2013).

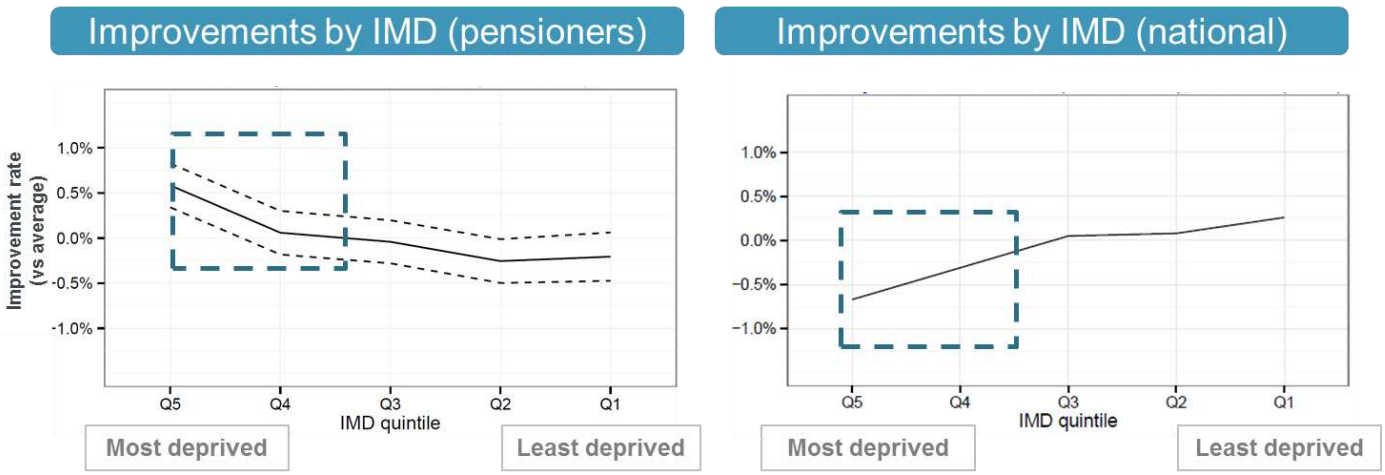


Figure 2.8: Average annual improvement rates by deprivation index (IMD 2007) for pensioner men using data from 1993 to 2011 (left hand side) and national data for men from 1995 – 2005, Lu et.al (2013) (right hand side)

At first sight the results look contradictory in that for the pension scheme data we see high improvements for the most deprived areas whereas research on the national data has found low improvements for the most deprived areas. The results for the least deprived areas are far more similar. One possible explanation for this 'deprivation paradox' is the difference in analyses i.e. multivariate vs univariate analysis. For example the univariate analysis of Lu may be confounded by an affluence effect associated with decreasing deprivation. However, we believe that we can largely rule this out as the adjustments in the GLM to allow for affluence (pension amount) are modest (see section 2.2.2.4). Thus this appears to be something specific to the nature of the two populations and potentially worthy of further research[16].

We believe that this feature is likely to be a consequence of a selection effect when focusing on pension scheme annuitants only. In the national data, those living in the most deprived areas will have, for example, high levels of unemployment and long term sickness. Those living in the most deprived areas but in occupational pension schemes are less likely to be typical of those areas. For instance, they are very unlikely to be long term unemployed. Indeed they may be those individuals improving their health outcomes via an element of upward socio-economic migration. So the annuitant data will be a very select – and different - subset of the national data. Such effects are often seen in the demographic literature where they are termed the 'ecological fallacy': a subgroup of individuals can exhibit very different pattern to the population as a whole (Greenland (2001)).

In short, conclusions drawn from the national data cannot be expected to translate well to the world of pension scheme and insurance company annuitants where there is a high level of socio-economic selection present. Thus care is needed in using models calibrated to *national* IMD data as they may be misleading in the management of *annuitant* basis risk.

[16] For example it could be insightful to look at additional variations of the multivariate model described here where multiple postcode rating factors (IMD and lifestyle) are included in the improvements component

2.2.2.4 Pension amount

Turning to the other key predictor of historical improvements, pension amount, the graphs in figure 2.9 demonstrates the relationship between affluence (current pension amount) and historical mortality improvement rates, controlling for the impact of deprivation shown in 2.1.2.2.

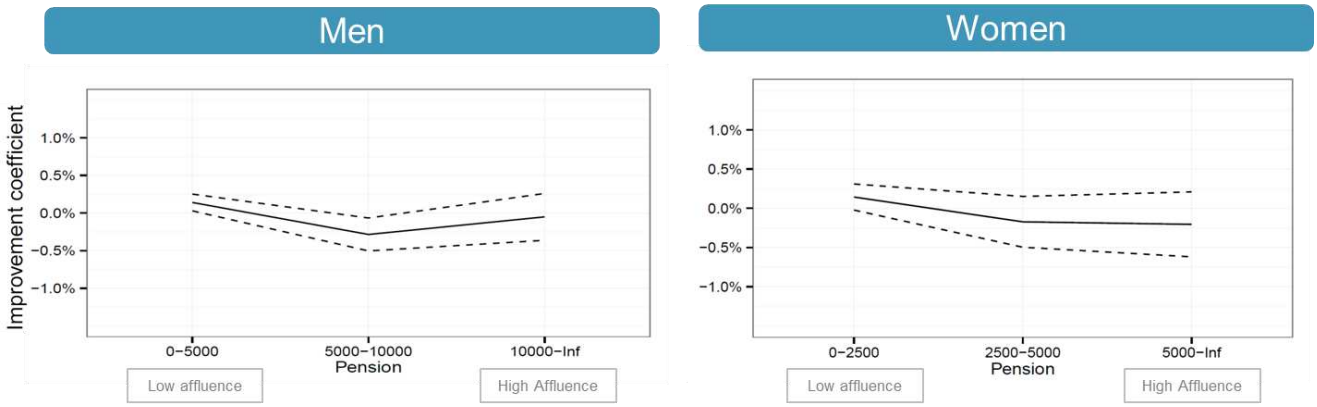


Figure 2.9: Pension amount component of annual improvements (relative to average level across all pension amounts)

The pension amount has been split into three pension bands – these are different for men and women, reflecting the lower pension amounts that have been accrued historically by women arising from differences in career and working patterns. For the avoidance of any doubt this analysis is restricted to pensioners only (i.e. excludes any in payment dependent pensions) in order to ensure comparability of pension amounts within each gender.

We can see how:

- the impact of affluence on mortality improvements is more modest than for deprivation, covering a spread of around 0.4% (figure 2.9) compared to around 0.8% for deprivation (figure 2.7).
- the differences in improvements between different affluence bands are generally weak (especially for women).
- the trend in improvements by affluence appears to have a 'smile' effect, especially for men, whereby higher improvements are being experienced by the lower and higher income pensioners.

2.2.3 Comparing different pension and deprivation combinations

Figure 2.10 demonstrates the materiality of differences in (fitted) annual improvement rates for a selection of different combinations of deprivation quintiles and affluence groups once compounded up over the period used to measure them (1993-2011).

In each case, the size of the bubble indicates the relative amount of pension in that group within Club Vita, and thus is a proxy to the financial significance of each group to the finances of a typical pension scheme. The number within the bubble is the level of total improvements between 1993 and 2011, with the red numbers reflecting above average increases.

Over this 18 year time period (1993-2011) the range in total mortality improvements has been between 36% and 49%. This 13% difference in mortality improvements is equivalent to a difference in liabilities of around 5%. This range clearly indicates the importance of demographic risk when assessing index-based solutions.

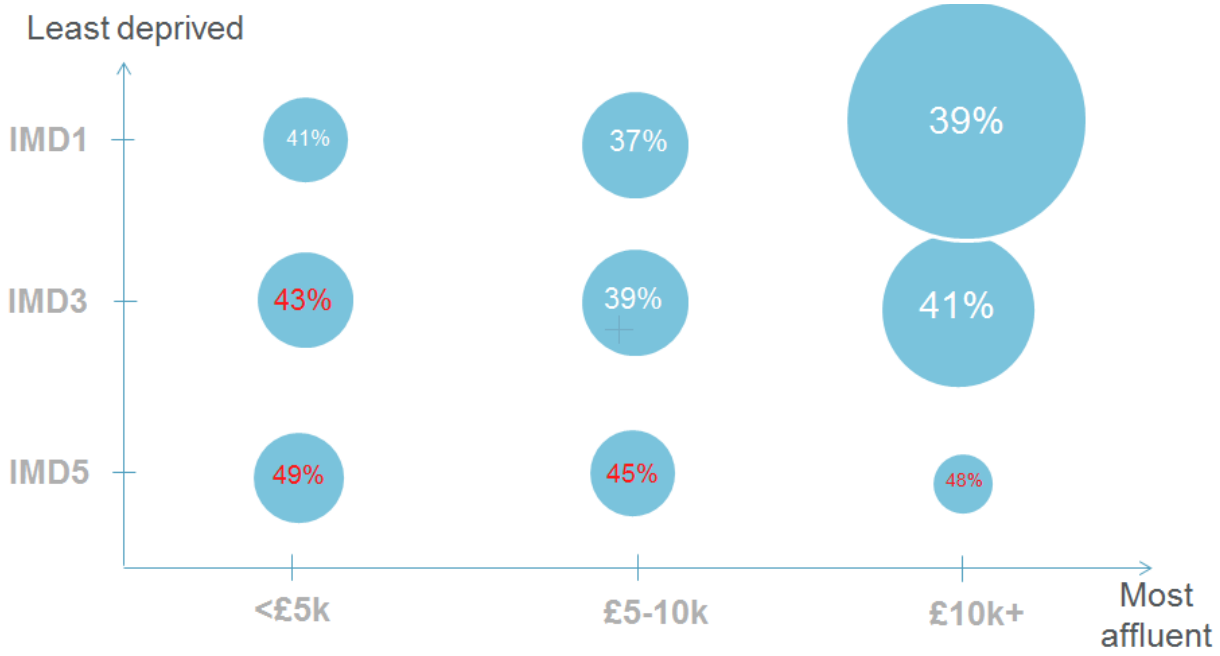


Figure 2.10: Total (fitted) improvement (reduction) in mortality rates (1993-2011) where differences in circle sizes refer to relative amount of pension for each socio-economic group.

2.3 Conclusions

The existing literature shows that material differences in mortality improvements have been identified, in particular when data has been segmented into groups according to gender, socio-economic class or deprivation.

Focusing on the Club Vita pensioner dataset, we observe that pension amount combined with deprivation had the strongest link to past improvements. Therefore, when looking for factors to characterise annuity data with respect to improvements, a combination of pension and deprivation is a good starting point. This result becomes very important when looking at the “characterisation approach” introduced in section 10.

We also observed how the IMD effect is very different in annuitant data to the whole population which is believed to be the consequence of a selection effect when focusing on annuitants only. Therefore, care is needed when using basis risk models parameterised using the whole UK IMD data.

Further, we have seen how the industry’s concerns around demographic risk being a significant issue is valid; differences in historical improvements by socio-economic classes have been of a similar magnitude to the gap between genders (which of the two is the only one currently allowed for when hedging longevity improvements).

3 High level review of drivers

We have shown that amongst those variables available to pension schemes two factors – deprivation as determined via postcode, and pension amount – are powerful in combination at modelling the historical observed improvements.

However, when considering the projection of mortality trends and how these might differ by socio-economic group it is important to understand the drivers of historical trends. This should inform matters of user judgement, such as the structural assumptions of any times series which drive modelling forecasts (see section 9.2.3).

3.1 The cohort effect

Much evidence has been presented for both the UK population and annuitants (both within the CMI dataset and the Club Vita dataset) experiencing a cohort effect. Specifically, the generation born broadly between the two World Wars are surviving in far greater numbers to an older age than their predecessors, as a consequence of experiencing markedly lower mortality rates than the generation preceding them. This is often illustrated using heat maps, such as the one in figure 3.1, which plot the annual reduction in mortality rates by age (age 0 to 100, y-axis) and calendar year (1962 to 2012, x-axis). Here the warm colours (yellows, oranges and reds) plot periods of particularly rapid reductions in mortality.

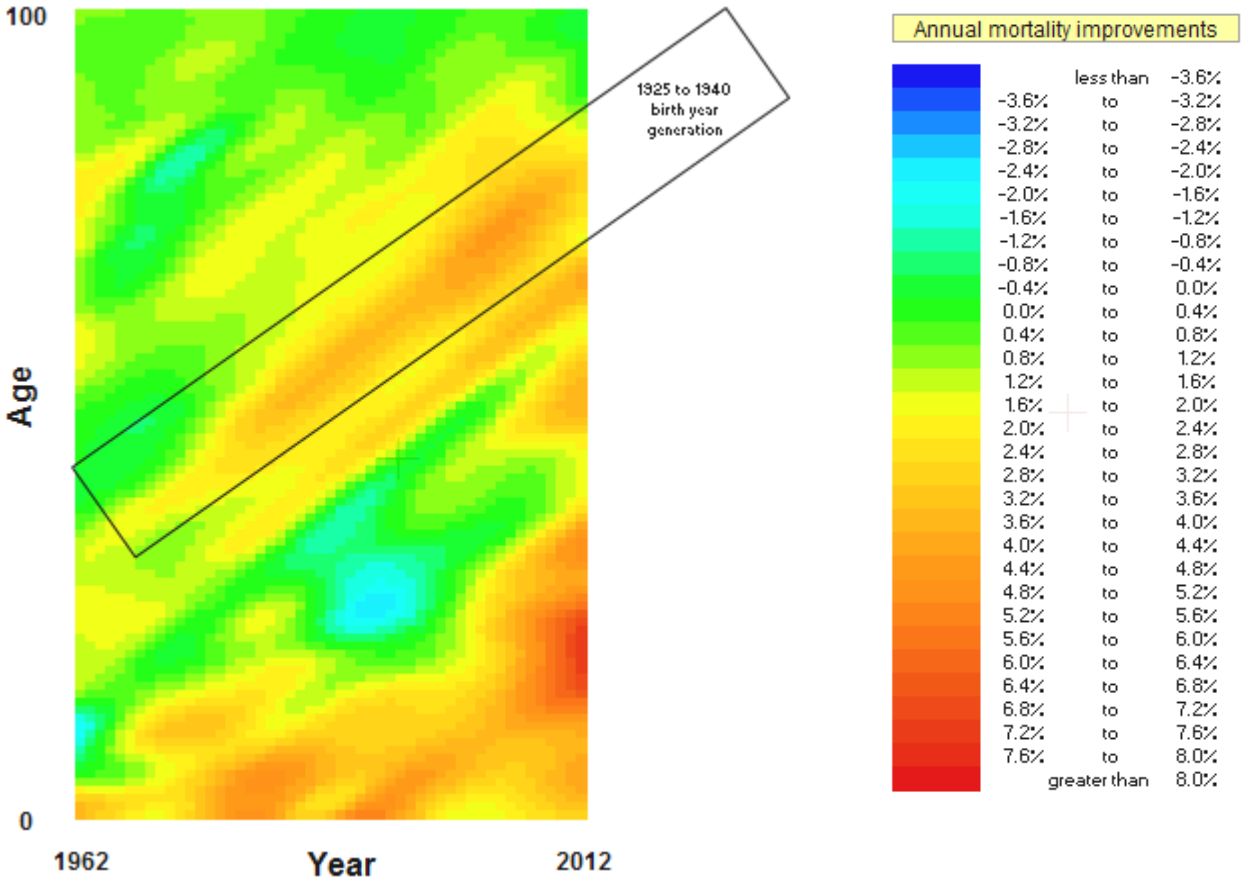


Figure 3.1: Heat map of p-spline improvements in England & Wales population data (ages 0-100, years 1962-2012). Age-cohort p-spline smoothing with 5 year knot spacings in both age and birth year dimensions. Underlying data as per that used in CMI 2013 sourced from http://www.actuaries.org.uk/research-and-resources/cmi-community/documents/cmi-mortality-projections-model-data-underlying-cmi20.

In charts such as this, birth generations move diagonally up the chart from bottom left to top right as they age. We can see a number of very clear cohort effects in this picture, including a period of strong improvements for those born between 1925 and 1940.

There are a number of reasons postulated for the cohort effect (see Willets (2004)) including:

- Introduction of the welfare state and particularly the NHS
- First generation to benefit from widespread use of antibiotics
- The positive impact conveyed by a high dietary intake of fresh vegetables and fish by the children growing up at this time
- Smoking cessation being most rapid amongst this generation

Whatever the underlying reasons, observations of this phenomenon led both GAD (see GAD 2001)) and the CMI to revisit previous projections. Within the context of modelling basis risk, it is important that our models allow for the now well-accepted cohort effect, separating out general improvements over time to those specific to a given birth cohort.

3.2 A causal perspective

Ultimately individuals die of something – low income or high deprivation per se do not kill individuals, although they will influence behaviours, the environment within which a person lives and their risk/predisposition to certain morbidities.

Figure 3.2 shows how age-standardised[17] mortality per 100,000 lives has been falling by the four major disease groups amongst the ages most relevant to demographic basis risk, ages 65+.

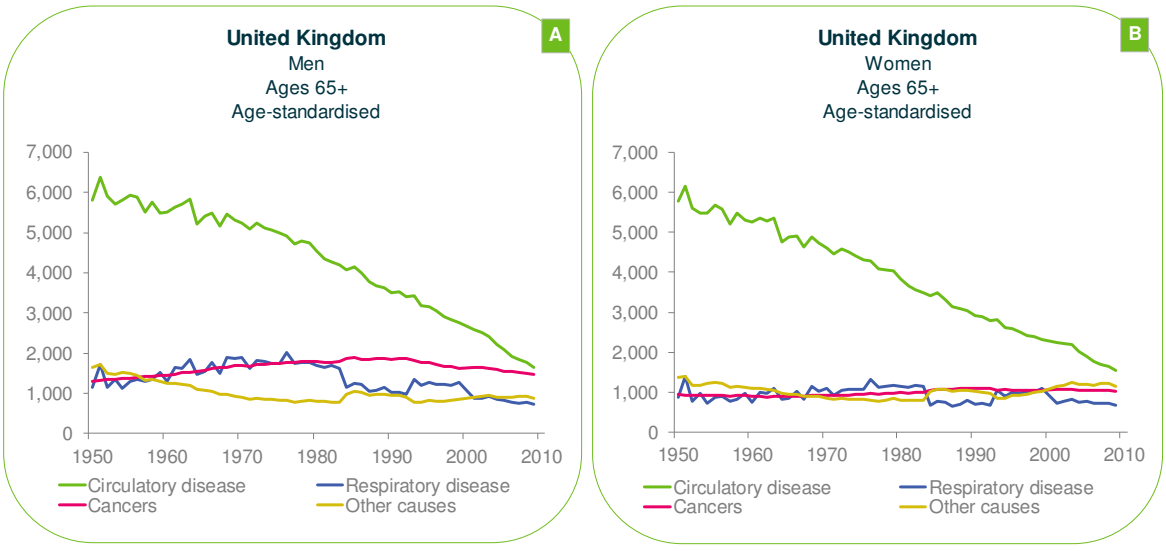


Figure 3.2: Age standardised mortality amongst UK over 65 year olds, per 100,000 lives (own calculations based on WHO data)

We can see how recent declines in mortality amongst over 65 year olds have been driven (within the UK) by dramatic declines in circulatory disease. Similar declines are seen in many other developed countries.

[17] Age standardised against 2008 UK population (5 year age grouped, ONS 2008 central projections) and using cause of death data sourced from WHO

In understanding the potential for continued falls in mortality it is helpful to consider what has driven the falls in circulatory diseases. In particular Belgin Unal and colleagues looked at changes in coronary heart disease (CHD) mortality amongst adults[18] in England & Wales between 1981 and 2000 (Unal et al (2004)). They identified that lifestyle and behavioural changes had been the biggest contributor to the decline - at 58% compared to 42% from medical interventions (figure 3.3).

Within the lifestyle and behavioural factors smoking dominates, accounting for nearly half the overall decline in heart disease mortality. The decline in smoking has, however, been different across the different professions and socio-economic groups.

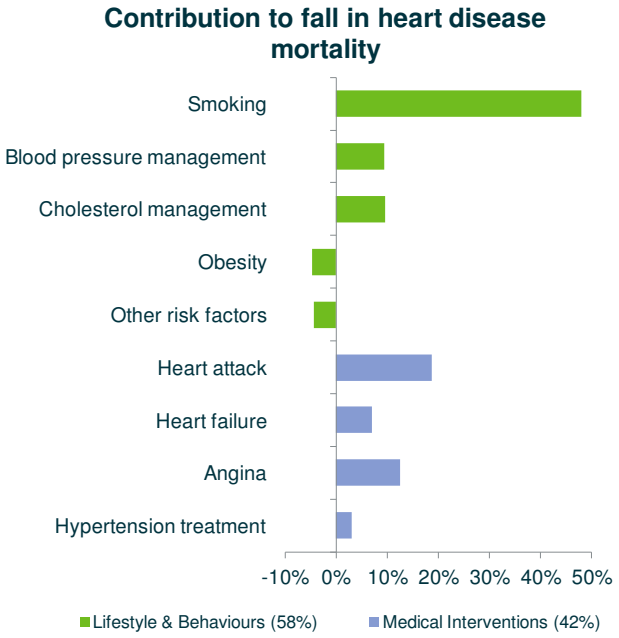


Figure 3.3: Attribution of change in CHD mortality amongst adults in England & Wales (summary of results of Unal et (2004))

For example ONS(2011) highlights how the proportion of smokers amongst manual occupations fell by around 25%[19] between 1992 and 2009, compared to over a 30% fall amongst non-manual occupations.

3.3 Differences within society

Different parts of society respond to the drivers of improving mortality in different ways. In the case of medical interventions for example there is some evidence that lower socio-economic groups are more reluctant to avail themselves of available resources (see for example Goddard & Smith (2001) and Morris et al (2005)), are less likely to be referred to specialist services (Dixon et al (2007)[20]), and have poorer adherence to treatment programmes, including for example the taking of regular medication (WHO (2003)).

With regard to lifestyle and behavioural factors one school of social epidemiology describes a 'social cascade' whereby the more educated socio-economic groups tend to be earlier, and fuller, adopters of healthier behaviours / new services such as the NHS. The same theory suggests that the less educated parts of society will be more sceptical and so be later, less whole-hearted adopters, tending to wait until they can see the positive effects in others.

[18] This study included individuals aged 25 to 84 - however as the majority of heart disease deaths occur in later life the results of that study should be relevant when considering trends amongst pensioner populations

[19] from 33% to 25%

[20] Note that Dixon et al (and Banks et al (2006)) also suggest that there is little evidence of socio-economic differences in accessing primary care. For a fuller discussion on these issues see LSE for the Equality and Human Rights Commission

Within the context of the drivers of UK mortality there is some limited empirical evidence to support this theory – for example the analysis of Evandrou & Falkingham (2002) of smoking patterns by manual and non-manual socio-economic group for a variety of birth cohorts.

However, one thing which can be stated with confidence is that the different socio-economic groups have different propensities to the different causes of death. We can see this in figure 3.4 which highlights how cancers (neoplasms) are a larger contributor to mortality amongst those living in the least deprived areas than to those living in the most deprived areas. Although the overall proportion of mortality associated with circulatory diseases is similar between these two groups, the overall mortality rate associated with each cause is noticeably higher for the most deprived group. In the case of circulatory disease this is likely to reflect the persistent socio-economic inequalities in risk factors such as obesity, systolic blood pressure and physical activity as highlighted by Scholes et al (2012).

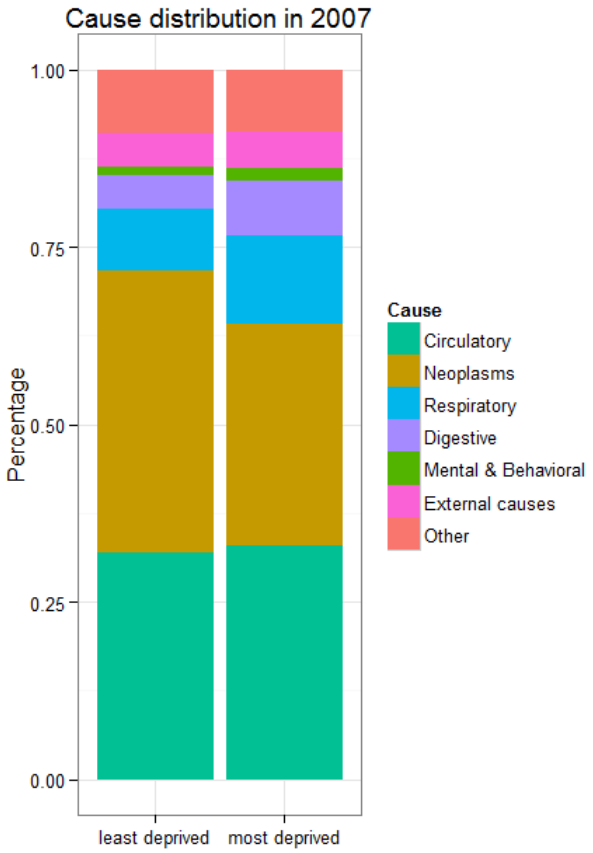


Figure 3.4: Split of mortality by cause of death for the most and least deprived quintiles of England, ages 25-84 (Villegas et al (2014))

To the extent these socio-economic groups will respond differently to government interventions, or be the focus of targeted health policies, demographic basis risk emerges.

3.4 Is the past a guide to the future?

Whilst we are able to cast some light on the drivers of historical trends, a natural concern is whether past improvements are a guide to future improvements. For example it is only possible to cease smoking once. It might therefore be argued that in the absence of replacement drivers future improvements might be slower than seen in the past, or indeed impinged by such factors as rising obesity (Olshansky et al (2005)). An alternative school of thought might be that historical trends are sustainable as medical and behavioural interventions successfully shift to keep pace with the prevailing causes of deaths of the time[21]. Naturally there is also the likelihood of new emerging drivers not yet identifiable in the historical data, and as the leading cause of death shifts from the circulatory disease to cancers the pace and shape of future improvements is changing. Caution is therefore needed in extrapolating the trends seen in historical data when modelling basis risk.

In the context of demographic basis risk we are particularly concerned with the scope for differences in the trends seen within the mix of lives within a specific book, and the reference population. In this regard the past has historically shown divergence amongst socio-economic groups (as per figure 2.2); however more recently, and particularly within the subset of lives likely to be covered by index swaps, we have seen faster

[21] See for example Baxter (2007) for further discussion of historical drivers and a selection of possible future drivers

improvements in the lower socio-economic groups (figure 2.10) and some convergence of mortality rates. When fitting models for the mortality of the book population vs the reference population care is therefore needed to:

1 Understand how the model reflects this shifting dynamic, and whether it implicitly incorporates an assumption of divergence or convergence of mortality and of life expectancies between the different socio-economic groups, or allows some possibility of either eventuality.

2 Interact with the modelling to apply user judgement where you have strong reason to believe that future outcomes may differ from those implied by historical trends.

4 Modelling problem

In order to be able to assess basis risk, we need a model that is able to capture the mortality trends in the reference population backing the hedging instrument and in the book population, the longevity risk of which is to be hedged. This modelling is needed in order to generate a distribution of future scenarios to evaluate the possibly different evolution of mortality rates in the two populations. Given this model, then the assessment of sampling risk and structuring risk becomes straightforward.

We denote by R the reference population and by B the book and assume that the following data is available.

- For the reference population,
  - D_{xt}^R number of deaths aged x last birthday in calendar year t
  - E_{xt}^R initial exposed to risk for age x and calendar year t.
  - The corresponding 1-year death rate for an individual in the reference population aged x last birthday and in calendar year t, denoted q_{xt}^R, can be computed as $q_{xt}^R = D_{xt}^R / E_{xt}^R$.
- Similarly, the corresponding quantities for the book population are denoted D_{xt}^B, E_{xt}^B and $q_{xt}^B = D_{xt}^B / E_{xt}^B$.

We assume that this data is available for a given set of ages and given numbers of years that can differ in the reference and the book. More precisely, we assume that D_{xt}^R, E_{xt}^R are available for consecutive ages $x = x_1, \dots, x_l$ and consecutive calendar years $t = t_1, \dots, t_{\mathrm{n_R}}$ in the reference population, while in the book D_{xt}^B, E_{xt}^B are available for ages $x = x_1, \dots, x_m$ and calendar years $t = u_1, \dots, u_{\mathrm{n_B}}$.

Typically, data for the reference population will be available over a longer horizon than in the book, that is $\mathrm{n_R} \geq \mathrm{n_B}$. Also, the set of calendar years of data in the book may be a subset of the corresponding calendar years in the reference population i.e. we may find that $u_{\mathrm{n_B}} \neq t_{\mathrm{n_R}}$. Further the ages available within the book may be a subset of those available in the reference population.

The modelling problem is then to identify a suitable model for q_{xt}^R and q_{xt}^B which produces consistent, stochastic, forecasts of future mortality[22].

[22] We would note that for user convenience we have chosen to work with one-year death probabilities $\boldsymbol{q_{xt}}$, as this a typical quantity of interest. However, if interested in central death rates, $\mathrm{m_x}$, or the force of mortality, μ_x, then the general modelling framework can be easily reformulated. When only central exposures are available and initial exposure are required, one can approximate the initial exposures to the risk of death by adding half the matching reported numbers of deaths to the central exposures (e.g. Section 2.2 Forfar et al., 1988). In addition, we do not expect to see any material differences in our analysis if central death rates, $\mathrm{m_x}$, or the force of mortality, μ_x, were considered instead.

5 Overview of available models

This section provides a general introduction to the available models to represent the mortality dynamics in the reference and the book populations. Figure 5.1 contains a schematic representation of the multi-population models currently available in the published literature, broadly grouped according to three main categories which we introduce in the next section.[23]

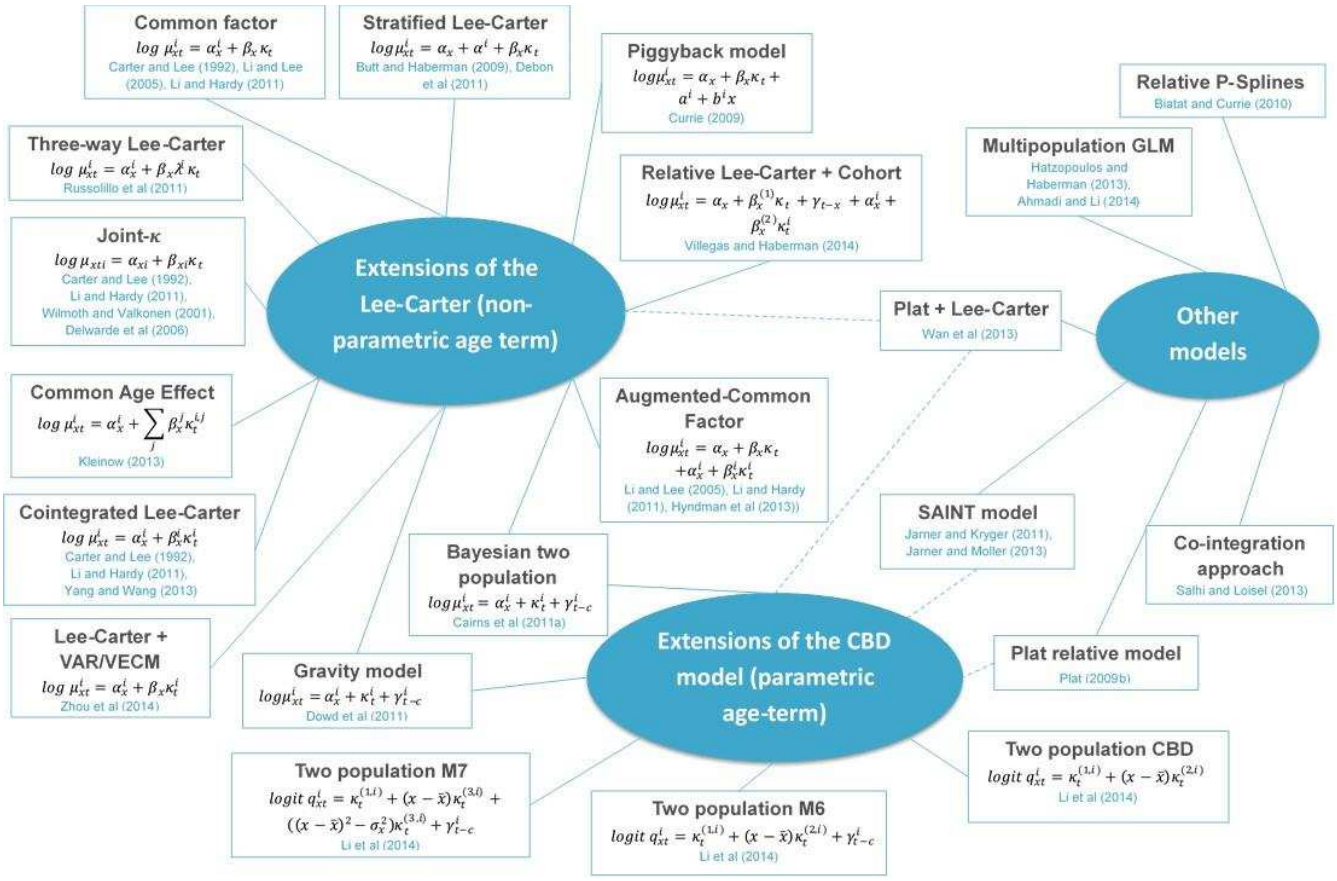


Figure 5.1: Universe of multi-population models

5.1 Literature review

Many models have been proposed in the literature to represent the mortality evolution of two or more related populations. All such contributions extend known single population models by specifying the correlation and interaction between the involved populations.

Although most of the academic contributions to the modelling of multi-populations are fairly recent, the first ideas go back to the seminal paper by Carter & Lee (1992), which suggested possible ways of extending their single population model in order to forecast differentials in US mortality between men and women.

Many existing models focus on the mortality rates of two or more related populations such as:

- National populations of different countries
- Men and women within a given country/population
- Smokers and non-smokers within a given country/population.

[23] Note that some models in Figure 5.1 are expressed in terms of force of mortality rather than the one year probability of death (our preferred choice)

A review and comparison of multi-population models can be found in Li and Hardy (2011), Villegas and Haberman (2014), Danesi et al. (2014), and Li et al. (2014).

Some relevant papers are as follows:

- Li and Lee (2005) first explicitly formulated the joint modelling of two related populations using an extension of the Lee-Carter model where specific and common period terms are included.
- Li and Hardy (2011) contains a comparison of some models in the context of assessing longevity basis risk.
- Cairns et al (2011b) and Jarner and Kryger (2011) recognize the relative importance of the reference population backing the index and the population whose longevity risk is to be hedged. Therefore, the model focuses on the reference population first and then on the spread between the reference and the book.
- Li et al. (2014) compare several two population extensions of the CBD – M5, M6 and M7 type models.

Many of the models shown in Figure 5.1, including both the extensions of the Lee-Carter approach (based on a non-parametric age term) and of the CBD approach (based on a parametric age term), can be fitted into a common framework – see the next section. There are, however, other contributions in the literature which attack modelling multi-population mortality from a different point of view. For instance, Biatat and Currie (2010) extend to two populations the P-spline methodology that has been successfully applied in the single population case, while Hatzopoulos and Haberman (2013) use a multivariate GLM.

5.2 Modelling the reference and the book population: A general formulation

We have identified a general framework under which most models that have been introduced in the literature can be accommodated. However, in order to facilitate this comparison between models, the way such models are proposed here may slightly differ from their original formulation.

As in Jarner and Kryger (2011) we choose a “relative approach” where the reference population is modelled first, and then the book mortality dynamics are specified so as to incorporate features from the reference population. This *relative* approach has some interesting features:

- It allows data mismatch between the reference and the book.
- It is well suited to the usual situation of the reference population being considerably larger than the book population.
- Reference population models are readily available and extensively studied, so this part of the model may be well established; allowing the focus to be on making an informed decision for the book part of the model, whilst retaining a good fit to the reference population.
- It provides consistency of approach when modelling several books using the same reference population.

Recall that D_{xt}^R and D_{xt}^B are the number of deaths aged x last birthday in the calendar year t in the reference population (R) and the book population (B) respectively. The corresponding initial exposures and 1-year death rates are E_{xt}^R, E_{xt}^B, q_{xt}^R and q_{xt}^B.

5.2.1 Reference population

A general model for the reference population can be written as[24]

$$D_{xt}^R \sim Bin(E_{xt}^R, q_{xt}^R) \quad (1)$$

$$logit(q_{xt}^R) = \log\left(\frac{q_{xt}^R}{1-q_{xt}^R}\right) = \alpha_x^R + \sum_{j=1}^{N} \beta_x^{(j,R)} \kappa_t^{(j,R)} + \gamma_c^R \quad (2)$$

Here:

- The term α_x^R determines the reference mortality *level* for age group x.
- N is some integer, allowing the user flexibility on the number of components which contribute to the mortality *trend* for the reference population with:
  - Each time index $\kappa_t^{(j,R)}$ contributing to the reference mortality *trend.*
  - Each coefficient $\beta_x^{(j,R)}$ dictates how mortality in the corresponding age group x reacts to a change in the corresponding time index $\kappa_t^{(j,R)}$ i.e. it modulates the sensitivity of the reference population at different ages to the general trend.
- The term γ_c^R is the cohort effect in the reference population (for birth cohort $c = t - x$).[25]

5.2.2 Book population

Given the reference population model, the book population is then specified through

$$D_{xt}^B \sim Bin(E_{xt}^B, q_{xt}^B) \quad (3)$$

$$logit(q_{xt}^B) - logit(q_{xt}^R) = \alpha_x^B + \sum_{j=1}^{M} \beta_x^{(j,B)} \kappa_t^{(j,B)} + \gamma_c^B \quad (4)$$

Note that we are modelling the difference in the (logit of) mortality in the book and the reference populations. Therefore:

- The term α_x^B determines the mortality *level* ***differences*** of the book population compared to the reference population for age group x. Hence the mortality level in the book is $\alpha_x^R + \alpha_x^B$.
- M is some integer (generally less than or equal to N), allowing the user flexibility on the number of components which contribute to the *trend* in differences in mortality between the book population and reference population with:
  - Each time index $\kappa_t^{(j,B)}$ contributes in shaping the difference in mortality *trends.*
  - Each coefficient $\beta_x^{(j,B)}$ dictates how mortality differences for age group x react to a change in the corresponding time index $\kappa_t^{(j,B)}$.
- The term γ_c^B accounts for the differences in cohort effect in the two populations (again for birth cohort $c = t - x$). Hence the cohort effect in the book is $\gamma_x^R + \gamma_x^B$

Depending on how the model is specified, identification constraints may have to be added to (1)-(4) in order to ensure that there is a single set of parameters which will yield a given set of mortality rates.

[24] Here, we have chosen to work with one-year death probabilities $\boldsymbol{q_{xt}}$. Therefore, it is most natural to use the logit function and model deaths using a binomial distribution. However, if interested in central death rates, m_x, or the force of mortality, μ_x, then the general modelling framework can be easily reformulated using a log link function and a Poisson Distribution.

[25] Note that equation (2) does not allow for an age-modulating factor in the cohort term. Models including such a factor have been considered by Haberman and Renshaw (2011) and Cairns et al (2009) and have been found to have robustness issues.

The estimation of the parameters of the model is performed in two stages whereby by the reference population part of the model is estimated in a first stage and then, conditional on the reference population parameters, the book population part of the model is estimated in a second stage[26].

5.2.3 Time series dynamics

The modelling is completed by specifying the dynamics of the period indices and the cohort terms which are needed for forecasting and simulating future mortality. Although alternatives have been explored by some authors (see e.g. Li et al (2013)) for the choice of the time series used in the dynamics, we (initially) confine our work to those commonly used in the literature. We discuss some of the consequences of this and alternative time series for the book population in section 9. A review of these basic time series choices and their properties is in Appendix C.

5.2.3.1 Reference population

Starting with the reference population, we assume that

$$\kappa_t^R = d + \kappa_{t-1}^R + \xi_t^R, \qquad d = \begin{bmatrix} d_1 \\ \vdots \\ d_N \end{bmatrix}, \qquad \kappa_t^R = \begin{bmatrix} \kappa_t^{(1,R)} \\ \vdots \\ \kappa_t^{(N,R)} \end{bmatrix}, \quad \xi_t^R \sim N(0, \Sigma^R) \tag{5}$$

$$\Delta\gamma_c^R = \phi_0 + \phi_1 \Delta\gamma_{c-1}^R + \varepsilon_c^R, \qquad \varepsilon_c^R \sim N(0, \sigma_R^2) \tag{6}$$

where Σ^R is an N x N variance-covariance matrix of the multivariate white noise ξ_t^R and $\Delta\gamma_c^R$ denotes $\gamma_c^R - \gamma_{c-1}^R$. Further:

- The time index κ_t^R is modelled as a multivariate random walk with drift (MRWD), so that a trend is implicitly assumed and the variance is growing with time (following Haberman & Renshaw (2011))
- The cohort index γ_c^R is modelled as an integrated auto-regressive process ARI(1,1) so as to capture a possible linear trend in the cohort effect when extending to the more recent birth years than covered by the data (following Renshaw & Haberman (2006))

5.2.3.2 Book population

As for the book population, we follow the assumption commonly made in the literature (see e.g. Jarner and Kryger (2011) and Li and Lee (2005)). More precisely we assume that in the long-run the two populations experience similar improvements[27] and therefore model the spread in the time indexes and in the cohort effects as stationary processes:

$$\kappa_t^B = \Phi_0 + \Phi_1 \kappa_{t-1}^B + \xi_t^B, \Phi_0 = \begin{bmatrix} \Phi_{01} \\ \vdots \\ \Phi_{0M} \end{bmatrix}, \quad \kappa_t^B = \begin{bmatrix} \kappa_t^{(1,B)} \\ \vdots \\ \kappa_t^{(M,B)} \end{bmatrix}, \quad \xi_t^B \sim N(0, \Sigma^B) \tag{7}$$

$$\gamma_c^B = \psi_0 + \psi_1 \gamma_{c-1}^B + \varepsilon_c^B, \qquad \varepsilon_c^B \sim N(0, \sigma_B^2). \tag{8}$$

where Σ^B is an M x M variance-covariance matrix of the multivariate white noise ξ_t^B and Φ_1 is an M x Mmatrix.

[26] An alternative approach would be to estimate simultaneously the parameters in the reference and book populations. This would in principle not materially change the fitted parameters as it is expected that book population has a small size relative to the reference population. A further possibility would be to fit the reference and book models in equations (1)-(4) jointly with the time series specified in (5)-(8) using a Bayesian approach as done in in Cairns et al (2011a).

[27] This will clearly have implications for the quantification of basis risk – a point we shall return to in section 9

- The time indices κ_t^B are then modelled as a vector auto-regressive process of order 1 (VAR(1)), for which we assume that the eigenvalues of the matrix Φ_1 are smaller than 1 in absolute value.
- The cohort difference γ_c^B follows an AR(1) process for which we assume that ψ_1 to be smaller than 1.
- We are assuming independence of the time series determining the reference population and those determining the difference between the reference and the book populations. Considering correlation between ξ_t^R and ξ_t^B or between ε_c^R and ε_c^B is in principle possible, as has been done e.g. in Cairns et al. (2011a). However, we have refrained from implementing this due to the fact that estimation of the correlations may become complicated. For example, this is the case when the time series for the reference and the book have different lengths, which is very frequent in practice.

5.3 Classification of models

The universe of two population mortality models covered by the general formulation in 5.2 can be broadly classified into the following, non-exclusive, categories:

- **Lee-Carter family (non-parametric age parameters)**: Here age is treated as factor, and the parameters α_x^i, $\beta_x^{(i,j)}$ $(i = R, B)$, when present, are not subject to any restriction and represent parameters to be estimated. An example is the augmented common factor model

$$logit(q_{xt}^R) = \alpha_x^R + \beta_x^R \kappa_t^R, \qquad logit(q_{xt}^B) - logit(q_{xt}^R) = \alpha_x^B + \beta_x^B \kappa_t^B$$

- **CBD family (parametric age structures):** Age is treated as a continuous variable and the parameters α_x^i, $\beta_x^{(i,j)}$ $(i = R, B)$ are specified as functions of age x and therefore do not need to be estimated. An example is the two population M5 model

$$logit(q_{xt}^R) = \kappa_t^{(1,R)} + (x - \overline{x})\kappa_t^{(2,R)}, \qquad logit(q_{xt}^B) - logit(q_{xt}^R) = \kappa_t^{(1,B)} + (x - \overline{x})\kappa_t^{(2,B)}$$

- **Other models:** Models that do not fit into the previous two families. For instance models that include both parametric and non-parametric age parameters such as the Plat+Lee-Carter model

$$logit(\, q_{xt}^R) = \, \alpha_x^R + \, \kappa_t^{(1,R)} + (x - \bar{x})\kappa_t^{(2,R)} + \gamma_{t-x}^R,$$

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

Figure 5.2 depicts the 3 categories of models and their relationship.

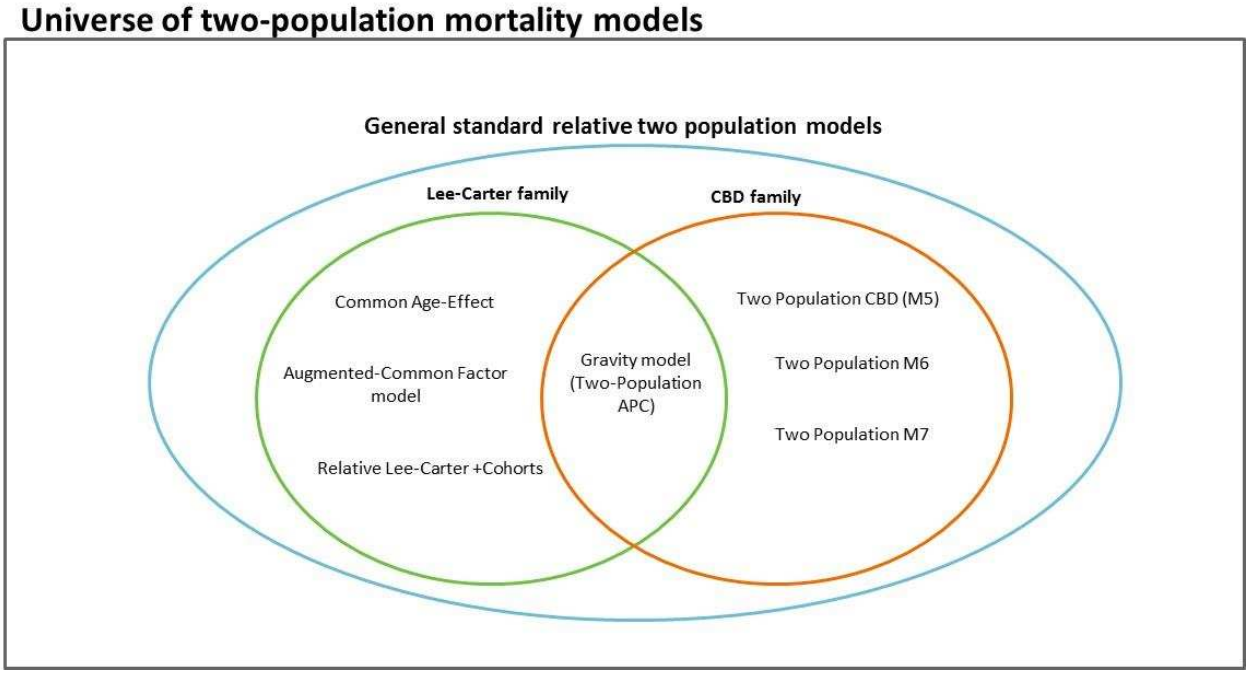


Figure 5.2. Universe of two-population mortality models

6 Identifying an appropriate two population model

With the vast number of existing two population models, the main problem is identifying which model(s) are most likely to provide a practical solution to assessing basis risk. To provide structure to this analysis, it is useful to test each model against certain criteria that a good and practical two population model for basis risk assessment should satisfy.

6.1 Criteria

Building on the literature comparing single population models (e.g. Continuous Mortality Investigation (2007); Cairns et al (2009); Cairns et al (2011b), Haberman and Renshaw (2011)), we consider the following criteria:

1 The model should be ***easy to implement*** using standard statistical methods likely to be available to practitioners.

2 The model should be ***transparent*** enough so that the model assumptions, limitations and outputs are understood by the users and can be easily explained to non-experts.

3 The model should be ***compatible*** with the data that is likely to be available when doing basis risk exercises.

4 The model should allow the ***disentanglement*** of level and improvement differences so that previous knowledge or alternative models for level differences can be readily incorporated.

5 The model should permit the consideration of a ***cohort effect*** if necessary.

6 The model should be relatively ***parsimonious***.

7 The model should produce a ***non-perfect correlation*** between year-on-year changes in mortality at different ages.[28]

8 The model should produce a ***non-perfect correlation*** between mortality rates in the two populations.[29]

9 The model should permit the ***generation of sample paths*** and the calculation of prediction intervals.

10 The structure of the model should allow the incorporation of ***parameter uncertainty*** *in simulations* using, for instance, bootstrapping techniques.

11 The model should show a reasonable ***goodness-of-fit*** to historical data in both the reference population and the book population for a wide range of book populations.

12 The model should show a reasonable ***goodness-of-fit*** for metrics involving the two populations such as differences or ratios in mortality rates or life expectancies for a wide range of book populations.

13 The model should produce mortality rates which are ***consistent*** with the observed and expected mortality characteristics e.g. be biologically reasonable with mortality increasing with age.

14 The model should produce ***plausible and reasonable best estimate projections*** of both single-population and two-population metrics.

[28] This refers to the correlation between $q^R_{x,t+1} - q^R_{x,t}$ and $q^R_{y,t+1} - q^R_{y,t}$ (or between $q^B_{x,t+1} - q^B_{x,t}$ and $q^B_{y,t+1} - q^B_{y,t}$) for $x \neq y$. Note that this correlation may not be perfect, although it will be close to one, even when correlation is perfect on logit scale used by the models introduced in section 5

[29] This refers to the correlation between q^R_{xt} and q^B_{zt}. Note that this correlation may not be perfect, although it will be close to one, even when correlation is perfect on the logit scale used for modelling.

15 The model should produce ***plausible and reasonable forecast*** level of uncertainty in projections of both single-population and of two-population metrics, which are in line with historical levels of variability.

16 The model should produce a ***non-trivial implied basis risk***.

17 The model should be ***robust*** to changes in the amount of exposures available in the book population.

18 The model should be ***robust*** to changes in the length of the historical data available in the book population.

19 The model should be ***robust*** to changes in the socio-economic composition of the book population.

Criteria 1 to 10 are theoretical properties of the model which can, in principle, be evaluated without reference to a specific data set, whereas criteria 11 to 19 can only be evaluated after the model has been fitted to data (and the conclusions drawn may therefore be dependent on the choice of dataset).

6.2 Assessment

Given the wealth of models available and the large number of criteria, we have followed a multi-stage filtering process to identify a shortlist of models likely to be suitable for basis risk assessment. In the first stage (section 6.2.1) we focus on data-independent criteria (criteria 1-10) and in the second and third stages we focus on those criteria that require data to be assessed (section 6.2.2 and section 6.2.3). More specifically, in the second stage of filtering we evaluate[30] the goodness-of-fit and the reasonableness of the output of various models (criteria 11-16), whilst in the third stage we investigate the robustness of those models which have passed the previous levels of filtering in order to ensure that they perform well in a wide range of circumstances (criteria 17 to 19). Appendix B contains further details of the assessment of each model against the criteria.

6.2.1 Stage 1 filtering: Criteria requiring no data to assess

We first evaluate all the candidate models against those criteria that can be assessed independently of data or the actual fitting of the models. This process permits the identification of a number of models which could be rejected, either because their theoretical properties are not suitable for basis risk assessment or because they are unlikely to be accessible to the wider industry. The main considerations which lead to models being rejected are:

- **Criterion 8 - Non-perfect correlation between mortality rates in the two populations:** If a model assumes or implies a *perfect correlation between mortality rates in the two populations* then it will imply that the reference population provides a perfect match for the book population. Although this might not be an issue for other purposes, this is clearly inappropriate for basis risk assessment as it will trivially lead to no (or very little) demographic basis risk[31]. This leads to the rejection of those Lee-Carter based models with a single common period effect for both populations including the Stratified Lee-Carter, the Piggyback Model, the Common Factor Model, the Three-way Lee-Carter, and the Joint-κ model.
- **Criterion 3 - Compatibility with available data:** The data requirements of some of the models are *incompatible with the likely available data*. For instance, it is unlikely that the book population will provide the same length of history as the reference population, hindering the application of models which cannot deal with such a scenario. In particular, this requirement leads to the rejection of two further Lee-Carter based models, namely the Lee-Carter VAR/VECM and the Co-integrated Lee-Carter. In addition, models

[30] As part of our assessment of the models we verified criteria 16 at the second stage, however we present the results verifying this as part of our case study

[31] A perfect correlation between the reference and the book populations always implies no or very little basis risk as the two populations move in parallel. However, depending on the basis risk metrics used, no or very little basis risk could occur for models where the two populations are not perfectly correlated. For instance, this is the case of the Relative Lee-Carter +Cohorts models when considering aggregate measures such as survival probabilities (see section 6.2.2.4 and Figures 6.6 and 6.8)

which use several book-specific period terms are poorly rated against the data compatibility criterion, as the more period terms a model has, the longer the data history that is required to estimate appropriately the time series processes needed for forecasting.

- **Criteria 1 & 2 – Ease of implementation and transparency:** *Ease of implementation* and *transparency* are essential for a model to be of general use by practitioners. Accordingly, these two criteria lead to the rejection of several other models. In particular, the Co-integration approach, the Relative P-splines and the Multipopulation GLM are considered to be impractical for basis risk assessment as they are complex models which are computationally involved to implement and may be difficult to communicate to non-experts. In addition, although the Bayesian Two Population model proposed in Cairns et al. (2011a) is particularly amenable to the short history and modest exposures sizes of most book datasets, the implementation and transparency issues related to the underlying Bayesian approach have led us not to consider this model. Finally, the Plat+Lee-Carter model was rejected (apart from other reasons discussed later) because it combines a parametric structure for the reference with a non-parametric structure for the book, and we believe that for the sake of interpretability of the parameters both parts of the model should be within the same family of models.

After carrying out this initial data-independent assessment 9 models were identified as candidates which are worth testing against the data dependent criteria (see figure 6.1). These models are: the Common Age Effect Model (with inclusion of a reference population cohort effect – see later), the Augmented-Common-Factor model, the Relative Lee-Carter models with cohorts, the Gravity model, the two-population M5, the two-population M6, the two-population M7, the SAINT model, and the Plat relative model.

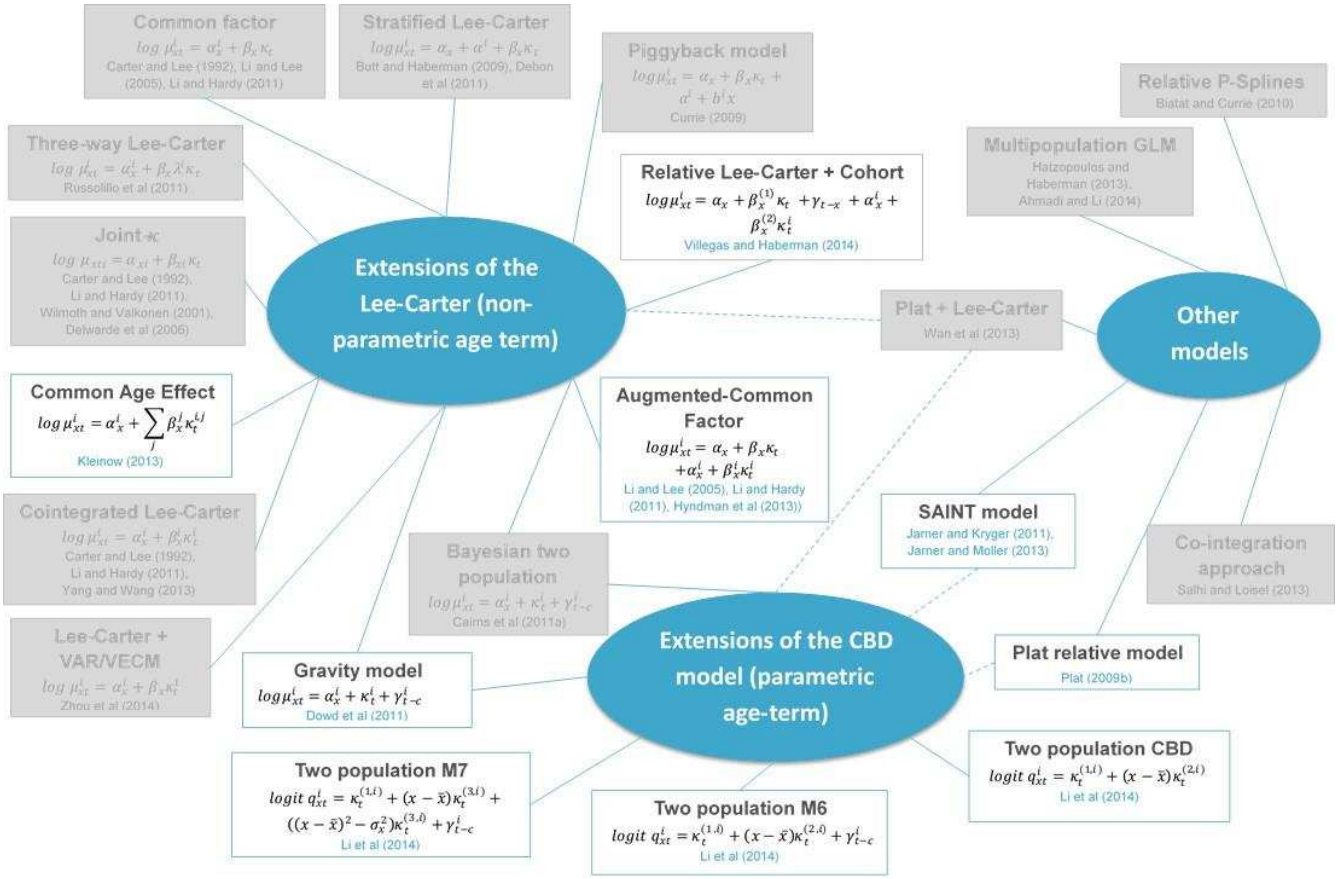


Figure 6.1: Remaining models after stage 1 filtering. Rejected models are greyed-out.

6.2.2 Stage 2 filtering: Reasonableness criteria requiring data to assess

In the second stage of filtering we focus on the reasonableness of fitting and output of the models. This involves the evaluation of the historical goodness-of-fit and the (subjective) evaluation of the reasonableness of the forecast level of uncertainty produced by the models. This stage enables us to further refine the list of candidate models before carrying out in stage 3 additional analysis on the robustness of a shortlist of candidate models.

6.2.2.1 Data

The evaluation of the reasonableness criteria requires data for model fitting. We have used as the reference population data the England and Wales male mortality experience as obtained from the Human Mortality Database (2013). For the purposes of our analysis we have focussed on a subset of this data covering calendar years 1961–2010 and those older ages as most relevant to longevity hedging, namely ages 60-89. [32]

For the book population we use synthetic datasets generated based on the profile of membership within individual Club Vita schemes but using the national mortality data split by IMD. The synthetic datasets used throughout this project have been generated by randomly sampling from the national data to obtain a dataset of exposure size, history length, and IMD profile desired. The technical details of this data sampling process are described in Appendix E.

The use of synthetic data as opposed to actual pension scheme data from Club Vita facilitates a more thorough assessment of the models. Concretely, synthetic datasets permit us to control some key characteristics of the book population data while changing others. For instance, it allows us to vary the history length and exposure size of the book data whilst keeping the socio-economic and age composition constant (preventing distortions arising either from gentrification or ageing of the portfolio over time). Moreover, synthetic datasets let us rely on the longer history of the national IMD mortality data to perform back testing exercises such us those described in Section 6.2.3.

For the assessment of the goodness-of-fit of the models, we consider four different synthetic datasets to reflect the variety of socio-economic mixes observed in real pension schemes and annuity books. In each case, the socio-economic splits are informed by the profiles seen within the Club Vita dataset. Table 6.1 describes the socio-economic profiles of these datasets while Figure 6.2 depicts the ratio of the mortality in each of the four datasets to the mortality in England and Wales. In all cases we use sample books with historical exposures of 100,000 male lives per year, which we believe is the largest exposure any scheme or insurer is likely to have. We also assume that book data are available for the period 1981–2010 and ages 60 to 89. (We return to smaller book sizes and shorter periods of experience data in 6.2.3.)

From Figure 6.2 it is worth noting that:

- The ordering of the ratios in the four data sets is consistent with their socio-economic mixes, with the “Extreme Wealthy” dataset having below average mortality (ratio < 1) and the “Extreme-Deprived” dataset having above average mortality (ratio > 1).
- The mortality ratios for both the “Typical Lives” and “Typical Amounts” dataset are close to 1 reflecting the socio-economic mix of these datasets being close to the average in England and Wales.
- In all of the datasets the mortality ratios converge with rising age. This is consistent with the commonly reported decrease in socioeconomic mortality differences as people age (Hoffmann, 2005).

[32] Although data are available above age 90, we have decided not to use them as age at death is often misreported at these higher ages resulting in unreliable estimates of mortality rates.

- None of the datasets show any very clear increasing or decreasing time trend in the mortality ratios, albeit there is a slight upward trend to the “Extreme Deprived” and a slight downward trend in the “Extreme Wealthy”

Dataset	Description	Percentage of exposure by IMD quintile				
		Q1	Q2	Q3	Q4	Q5
Typical Lives	This is the typical IMD split we would expect to see in a book population weighted by lives (head-count).	23%	22%	21%	20%	14%
Typical Amounts	This uses the same split as the typical (lives) but weighted by individual pension amounts to approximate the effect of a typical portfolio’s liability distribution amongst the IMDs	30%	25%	20%	15%	10%
Extreme Wealthy	This reflects the split by IMD (on an amounts weighted basis) that we would expect to see in a very affluent book population	45%	30%	20%	5%	0%
Extreme Deprived	This reflects the split by IMD (on a lives weighted basis) that we would expect to see in a book skewed towards lower socio-economic groups	10%	15%	15%	25%	35%

Table 6.1: Description of the book datasets used for model testing

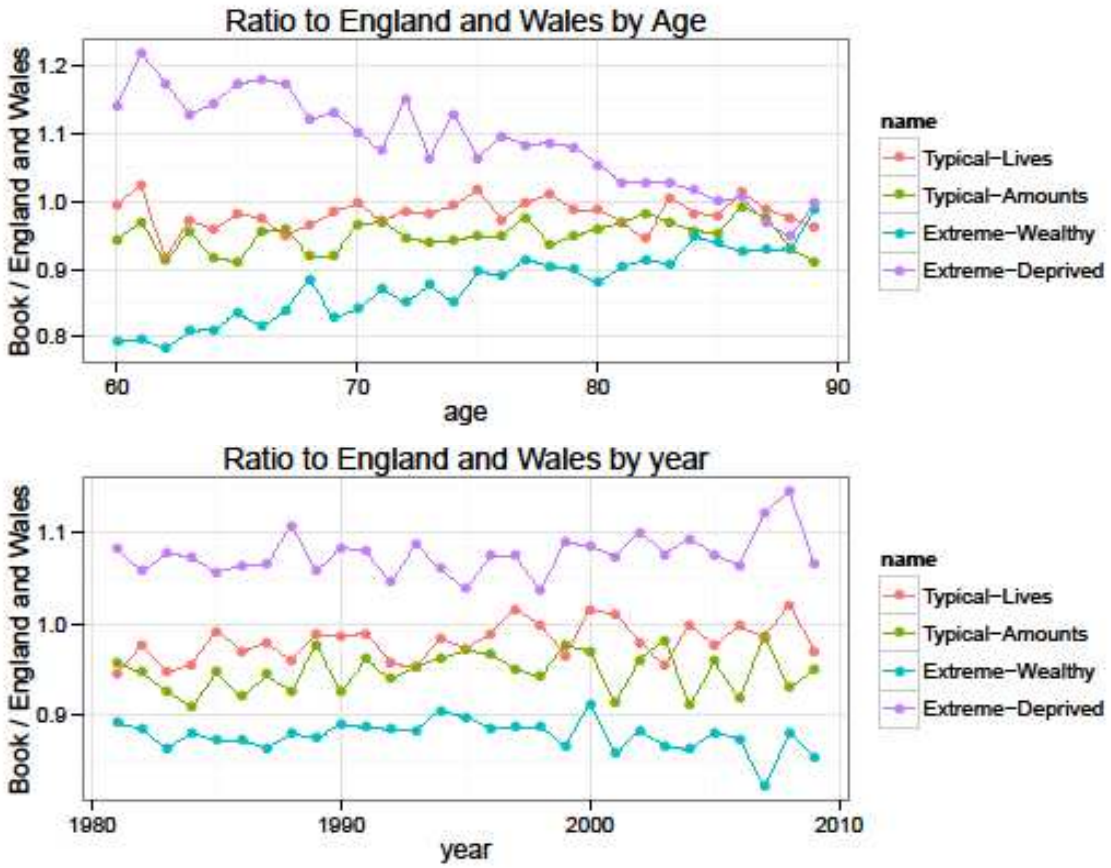


Figure 6.2: Ratio of the mortality in each of the four synthetic book datasets to the mortality in England and Wales. The top graph shows this ratio by age while the bottom one presents the time evolution of this ratio

6.2.2.2 Model fitting

To facilitate the fitting of the 9 models that passed our first-stage filtering, we have followed the general modelling framework described in section 5.2 whereby each model can be viewed as a model for the reference population combined with a model for the book population (or perhaps more accurately, a model for the logit difference between reference and book). As such, the fitting and the assessment of the goodness-of-fit of a model can be carried out in two stages: fitting and assessing the goodness-of-fit of the reference model, followed by the fitting and the assessment of the goodness-of-fit of the book part of the model.

We note that conclusions regarding the goodness-of-fit of the model to the reference may lead us to slightly modify the original formulation of certain of the two-population models before assessing the goodness-of-fit of the book part of the model. The specific modifications for each particular two-population model are described later in this section.

6.2.2.3 Goodness-of-fit for the reference population

For the reference population we concentrate on the six models described in Table 6.2 which are labelled Lee-Carter, Lee-Carter+Cohorts, APC, M5, M6 and M7. The Lee-Carter+Cohorts[33] is one of the Renshaw & Haberman (2006) extensions of the original Lee-Carter model. The APC model is a special case of the Lee-Carter+Cohorts. Models M6 and M7 are extensions of the original CBD model (M5) and were proposed in Cairns et al. (2009). These models encompass the reference population models underlying the models that passed our first stage filtering with the exception of the SAINT model. For the SAINT model, instead of the frailty-type model considered originally by Jarner & Kryger (2011) which we believe is too complex to be accessible to practitioners, we will use M7 to model the mortality of the England and Wales reference population.

Model	Formula
Lee-Carter	$\text{logit}(q^R_{xt}) = \alpha^R_x + \beta^R_x \kappa^R_t$
Lee-Carter + Cohorts	$\text{logit}(q^R_{xt}) = \alpha^R_x + \beta^R_x \kappa^R_t + \gamma^R_{t-x}$
APC	$\text{logit}(q^R_{xt}) = \alpha^R_x + \kappa^R_t + \gamma^R_{t-x}$
M5	$\text{logit}(q^R_{xt}) = \kappa^{(1,R)}_t + (x - \bar{x})\kappa^{(2,R)}_t$
M6	$\text{logit}(q^R_{xt}) = \kappa^{(1,R)}_t + (x - \bar{x})\kappa^{(2,R)}_t + \gamma^R_{t-x}$
M7	$\text{logit}(q^R_{xt}) = \kappa^{(1,R)}_t + (x - \bar{x})\kappa^{(2,R)}_t + ((x - \bar{x})^2 - \sigma^2_x)\kappa^{(3,R)}_t + \gamma^R_{t-x}$

Table 6.2: Mathematical description of the six models considered for the reference population.

We first assess the goodness-of-fit of the candidate reference population models by examining sign plots of deviance residuals. Regular patterns in the residuals are an indication of the inability of the model to describe all of the features of the data appropriately. Figure 6.3 plots the sign of the residuals in an age-period grid for the six reference population models. From this figure we note the following:

[33] It is well known that cohort extensions of the Lee-Carter model have robustness and stability issues (see e.g. Cairns et al. (2009)) with model being very sensitive to changes in the data or the fitting algorithm. Therefore, when implementing the Lee-Carter+Cohorts model we do not consider an age-modulating factor in the cohort term and follow the approach suggested in Hunt & Villegas (2014) which helps resolve many of the stability issues. (See page 59 for further details).

- The Lee-Carter and M5, which do not incorporate a cohort effect, show diagonal clusters of positive and negative residuals. This provides strong evidence for the existence of a cohort effect in the England and Wales reference population.
- The APC models show a strong clustering of positive and negative residuals. This is due to its inability to allow for varying improvement rates with age.

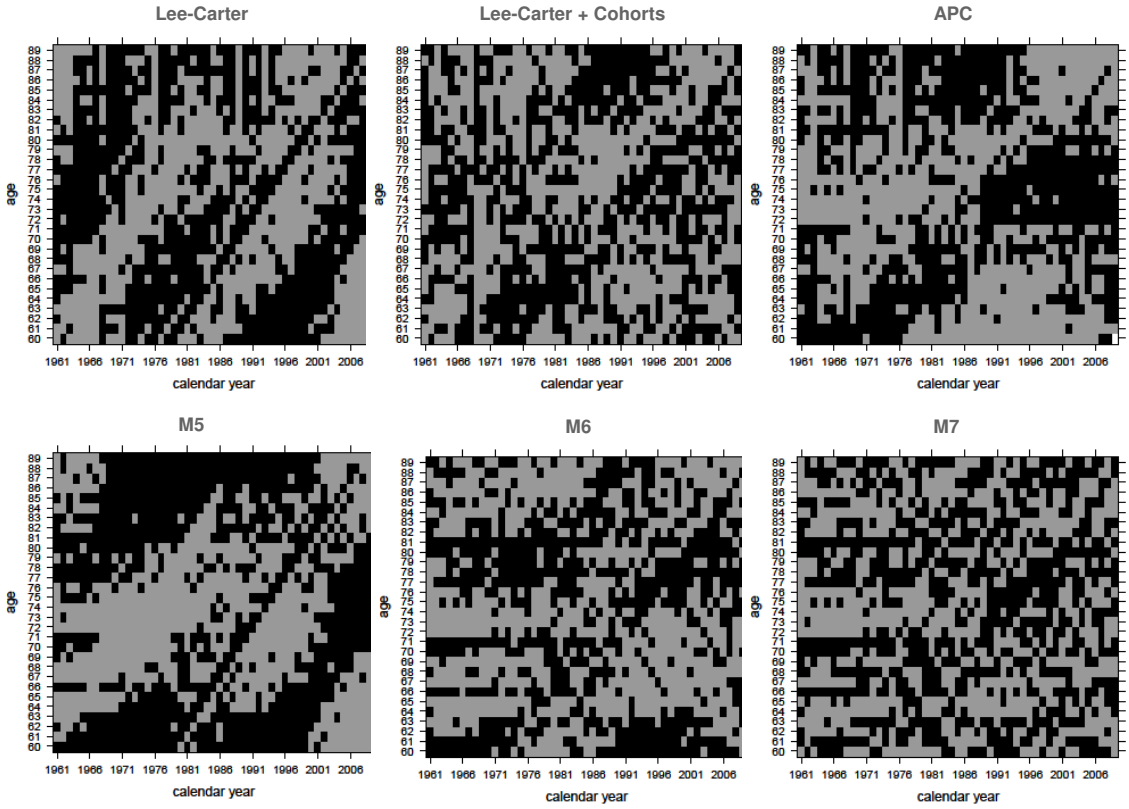


Figure 6.3: Sign plots of deviance residuals for the England and Wales males reference population. Positive residuals in grey and negative residuals in black.

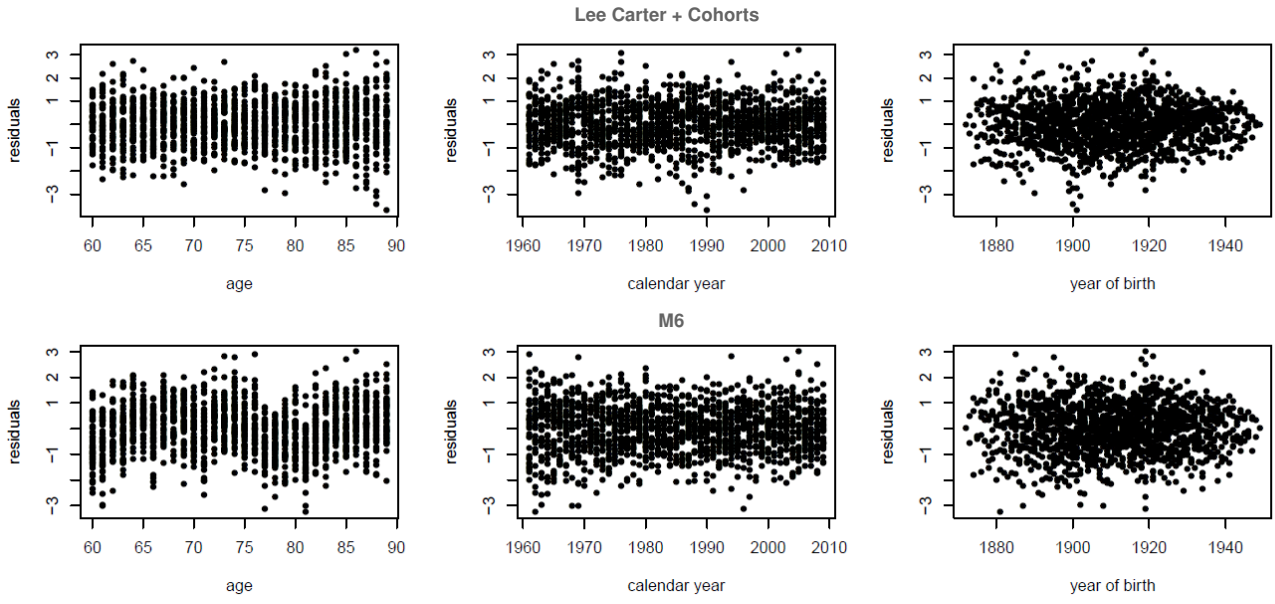


Figure 6.4: Scatterplots of deviance residuals for the England and Wales males reference population using models Lee-Carter+Cohorts and M6.

- The Lee-Carter+Cohorts and M6 look reasonably random, but still with some clustering of positive and negative residuals. Closer inspection of scatterplot of the residuals of these two models (see Figure 6.4) reveals that the Lee-Carter+Cohorts doesn't show any clear pattern while model M6 shows a strong pattern by age. This latter pattern reflects the lack of a quadratic age term in model M6 which may be necessary to capture the commonly observed curvature of the mortality rates in a logit scale.
- The M7 look reasonably random, indicating a satisfactory fit to the data.

In order to assess more formally the goodness-of-fit of the six reference population models, we present in table 6.3 the Akaike Information Criterion (AIC), defined as $2\nu_R - 2\ell_R$, where ℓ_R is the loglikelihood of the reference model and ν_R is the number of parameters of the models. The AIC provides a way of assessing the balance between quality of fit and parsimony. In general, a lower value of AIC is preferable. We note that the Lee-Carter+Cohorts and M7 are the best fitting models.

Model	Number of parameters	AIC (rank)
Lee-Carter	107	77,500,234 (5)
Lee-Carter + Cohorts	183	77,494,610 (2)
APC	154	77,496,340 (4)
M5	98	77,501,279 (6)
M6	174	77,494,736 (3)
M7	222	77,494,283 (1)

Table 6.3: Effective number of parameters, and AIC for different models fitted to the England and Wales reference population. AIC rankings across models are presented in brackets

The previous observations are consistent with the existing literature comparing single population mortality models (e.g. Cairns et al. (2009) and Haberman & Renshaw (2011)), where the Lee-Carter+Cohorts and M7 have been identified as appropriate models for modelling mortality in the England and Wales population. Accordingly, in our subsequent evaluation of two-population extensions of the Lee-Carter model, we will assume that the reference population is modelled using a Lee-Carter model with cohorts. Similarly, when assessing the two-population extensions of the CBD model, we will assume that the reference population is modelled using an M7 model.

6.2.2.4 Goodness-of-fit for the book population and two population metrics

To correct some of the goodness-of-fit issues discussed above, we have adapted several of the candidate two-population models before carrying out further goodness-of-fit assessments. Specifically, we have made the following adaptations:

- The Common Age Effect model, as proposed in Kleinow (2013), does not include a cohort effect. Therefore, given that there is strong evidence of a cohort effect in England and Wales, in our testing we extend this model to include such an effect. The reference population model is then a Lee-Carter + Cohorts model in the terminology of 6.2.1.
- Similarly, for the Augmented Common-Factor model we should consider a cohort effect, but doing so would turn the model into the Relative Lee-Carter model with cohorts. Consequently, the Augmented Common-Factor model is not considered further in the analysis.
- In the two-population M5 and the two-population M6 models we replace the corresponding M5 and M6 models for the reference population with an M7 model.

- For the Relative Plat model we assume an M7 model for the reference population as opposed to the M5 model originally assumed by Plat (2009b).
- For the SAINT model we assume an M7 model for the reference population instead of the frailty-type model originally used by Jarner & Kryger (2011).

For comparison purposes, in some of our additional goodness-of-fit and reasonableness testing we will consider the Common Factor Model with cohorts. This model, which was previously deemed inappropriate as it unrealistically implies zero basis risk, is useful for illustrating some of the undesired characteristics in a model to put the other models into context.

Table 6.4 (next page) summarises the models whose goodness-of-fit will be investigated further. The Common Factor models with cohorts (CF+Cohorts), the Common Age Effect model with cohorts (CAE+Cohorts), and the relative Lee-Carter model with cohorts (RelLC+Cohorts) belong to the Lee-Carter family of models described in section 5.3. The CF+Cohorts only allows for level differences between the reference and the book population, whilst the CAE+Cohorts and the RelLC+Cohorts also allow for improvement differences. Nevertheless, the latter two models differ in the specification of the age-modulating factor β_x^B accompanying the book-specific time index κ_t^B: in the RelLC+Cohorts β_x^B is estimated directly from the observed logit difference of mortality between the book and reference data, whilst in the CAE+Cohorts β_x^B is inherited from the reference population model, i.e. $\beta_x^B \equiv \beta_x^R$.

Models M7-M5, M7-M6, M7-M7, M7-SAINT, and M7-Plat (which are the implemented versions of the two population CBD, the two population M6, the two population M7, the SAINT model, and the Relative Plat model, respectively) all belong to the CBD family of models. These models differ in the type of differences between the book and the reference population that are allowed for in the parametric age functions:

- M7-M5 and M7-M6 allow only for level and slope differences with M6 also allowing for cohort differences;
- M7-SAINT, M7-M7 allow for level, slope and curvature differences with M7 also allowing for cohort differences; and
- M7-PLAT is a constrained version of M7-M5 assuming that at age 100 there is no difference between the reference and the book.

Original model	Model tested	Reference formula $(\mathbf{logit}(q_{xt}^R))$	Book differences formula $(\mathbf{logit}(q_{xt}^B) - \mathbf{logit}(q_{xt}^R))$
Common Factor	CF+Cohorts	$\alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R$	α_x^B
Common Age Effect	CAE+Cohorts	$\alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \beta_x^R \kappa_t^B$
Relative Lee-Carter with cohorts	RelLC+Cohorts	$\alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \beta_x^B \kappa_t^B$
Gravity	Gravity (APC)	$\alpha_x^R + \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \kappa_t^B + \gamma_{t-x}^B$
Two-population M5	M7-M5	$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R$	$\kappa_t^{(1,B)} + (x-\bar{x})\kappa_t^{(2,B)}$
Two-population M6	M7-M6	$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R$	$\kappa_t^{(1,B)} + (x-\bar{x})\kappa_t^{(2,B)} + \gamma_{t-x}^B$
Two-population M7	M7-M7	$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R$	$\kappa_t^{(1,B)} + (x-\bar{x})\kappa_t^{(2,B)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,B)} + \gamma_{t-x}^B$
SAINT model	M7-SAINT	$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R$	$\kappa_t^{(1,B)} + (x-\bar{x})\kappa_t^{(2,B)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,B)}$
Plat relative model	M7-Plat	$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R$	$\frac{100-x}{100-\bar{x}}\kappa_t^{(1,B)}$

Table 6.4: Mathematical description of the two population models considered for goodness-of-fit assessment. In the equations $\bar{x}$ is the average age in the data and σ_x^2 is the average value of $(x-\bar{x})^2$. Note the commonality of the reference formula for the first three rows, and similarly for two–population M5 onwards.

A good two-population model should show a reasonable fit to the historical mortality rates in both the reference population and the book population. In addition, the model should show a good fit to metrics involving the two populations such as differences or ratios of mortality rates. This last criterion is very relevant as demographic basis risk emerges from the mismatch in the mortality of the reference and the book population.

When assessing the quality of the fit of the models with respect to the book population and with respect to two-population metrics, we have found that the inspection of residual sign plots is not very informative. In principle, this can be attributed to the fact that cohort and age patterns in the book residuals may be confounded with the sampling noise in the book data.

Indeed when assessing the goodness-of-fit to mortality at individual ages we run the risk of focussing on the goodness-of-fit at a specific age (typically less than 5% of data) and thus gearing up sampling noise. In practice users of the basis risk methodology will use q- or s- forward structures in order to provide a hedge against anticipated annuity payments. Consequently there is less interest in precisely hedging mortality at a single age, rather hedging the mortality dynamic over a range of ages. As an alternative we have therefore examined (for this phase of the work) metrics closely related to the quantities someone entering an index-based hedge will be seeking to hedge, annuity payments across the age spectrum. We therefore examine plots of fitted vs. observed period survival probabilities in the book and the corresponding plots for ratios of period survival probabilities in the book and the reference can give useful insight into the goodness-of-fit of the models.

As an illustration figure 6.5 depicts the fitted and observed 30 years period survival probabilities at age 60 for the 'Typical Lives' sample scheme using several models. Figure 6.6 plots the corresponding ratios of fitted to observed period survival probabilities between the 'Typical Lives book' and the England and Wales reference.

Figure 6.5 shows that, with the exception of the M7-Plat model which shows a slight underestimation in the later years, all the models show a similar and reasonable fit to the period survival probabilities in the book. By contrast, when considering ratios of survival probabilities the models show very different performances. In particular, from Figure 6.6 we note:

- The M7-Plat model shows some bias in the fitted ratios consistent with the underestimation seen in the period survival probabilities in the book population. The poor fit of this model becomes more evident when considering the 'Extreme Deprived' dataset (see Figure 6.7). This suggests that the M7-Plat model might be too restrictive for some datasets and, thus, we do not consider it further as a candidate for basis risk assessment.
- The CF+cohort and the RelLC +Cohort models produce very smooth ratios of survival probabilities which seem to understate the observed volatility in the ratios. Whilst the poor performance of the CF+cohort model was expected due to perfect correlation between populations assumed by this model, the poor performance of RelLC +Cohort was not.
- Further investigation of the parameters of the RelLC +Cohort, indicates that the over-smoothed fitted ratios can be linked to the presence of a book-specific non-parametric β_x^B which needs to be estimated from the book data. The estimation of this parameter requires large amounts of data, and, hence, with the relatively small population sizes of the book populations, the estimated β_x^B values tend to be erratic and lack robustness. In particular, there exists the possibility that β_x^B fluctuates around 0 (see Figure 6.8) which results in mortality differentials between the book and the reference cancelling out when aggregated measures of mortality such as survival probabilities and life expectancies are calculated. Given that this over fitting of the β_x^B may result in an inappropriate perfect correlation between the reference and the book populations, we consider that the RelLC +Cohort is inadequate for basis risk assessment. This conclusion extends to other models with non-parametric β_x^B parameters such as the Augmented Common-Factor model and the Plat+Lee-Carter model.

- Some models, such as M7-SAINT and M7-M7 show signs of potential over-fitting to the data

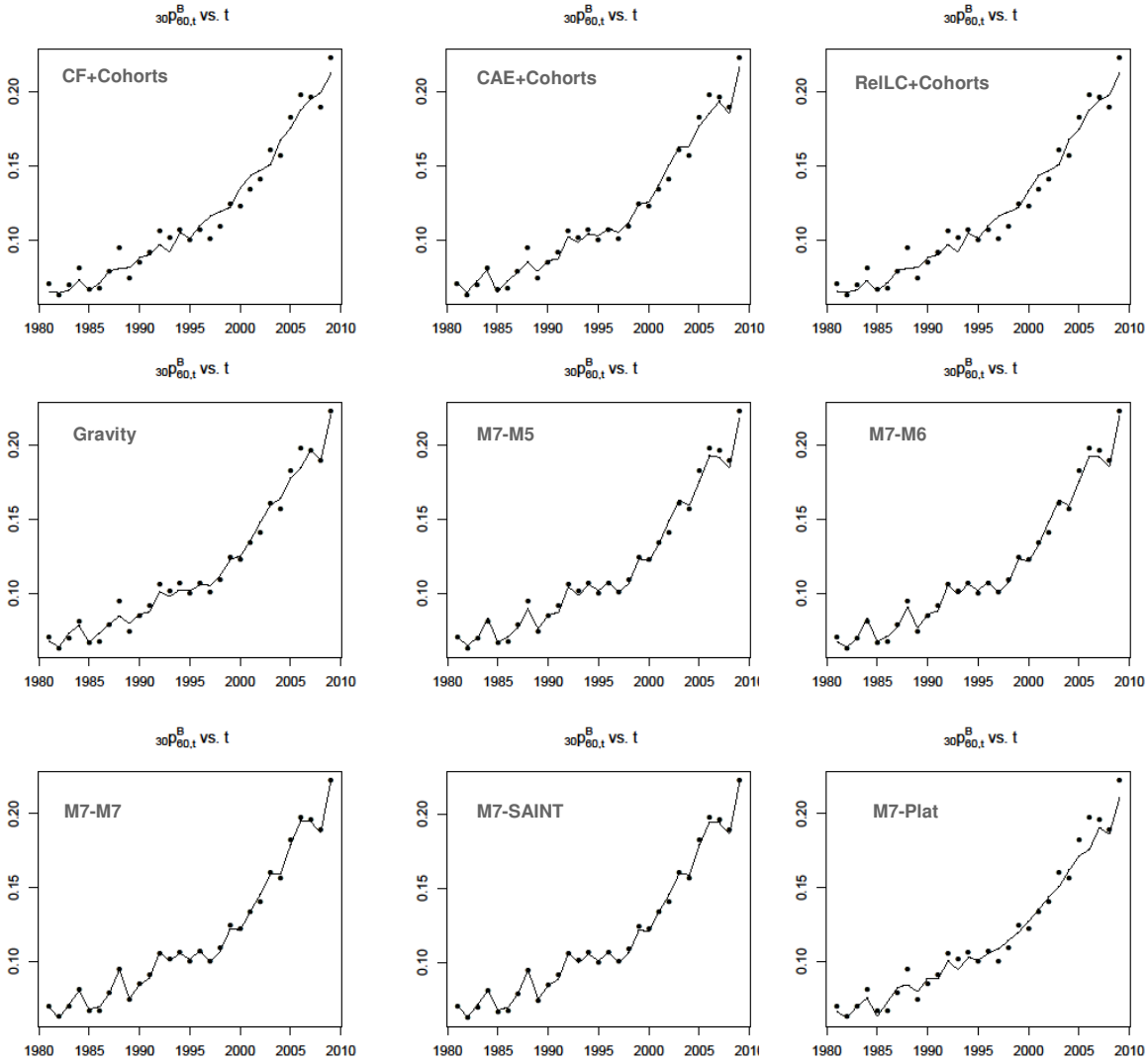


Figure 6.5: Fitted vs. observed 30 year period survival probabilities at age 60 for the 'Typical Lives' scheme

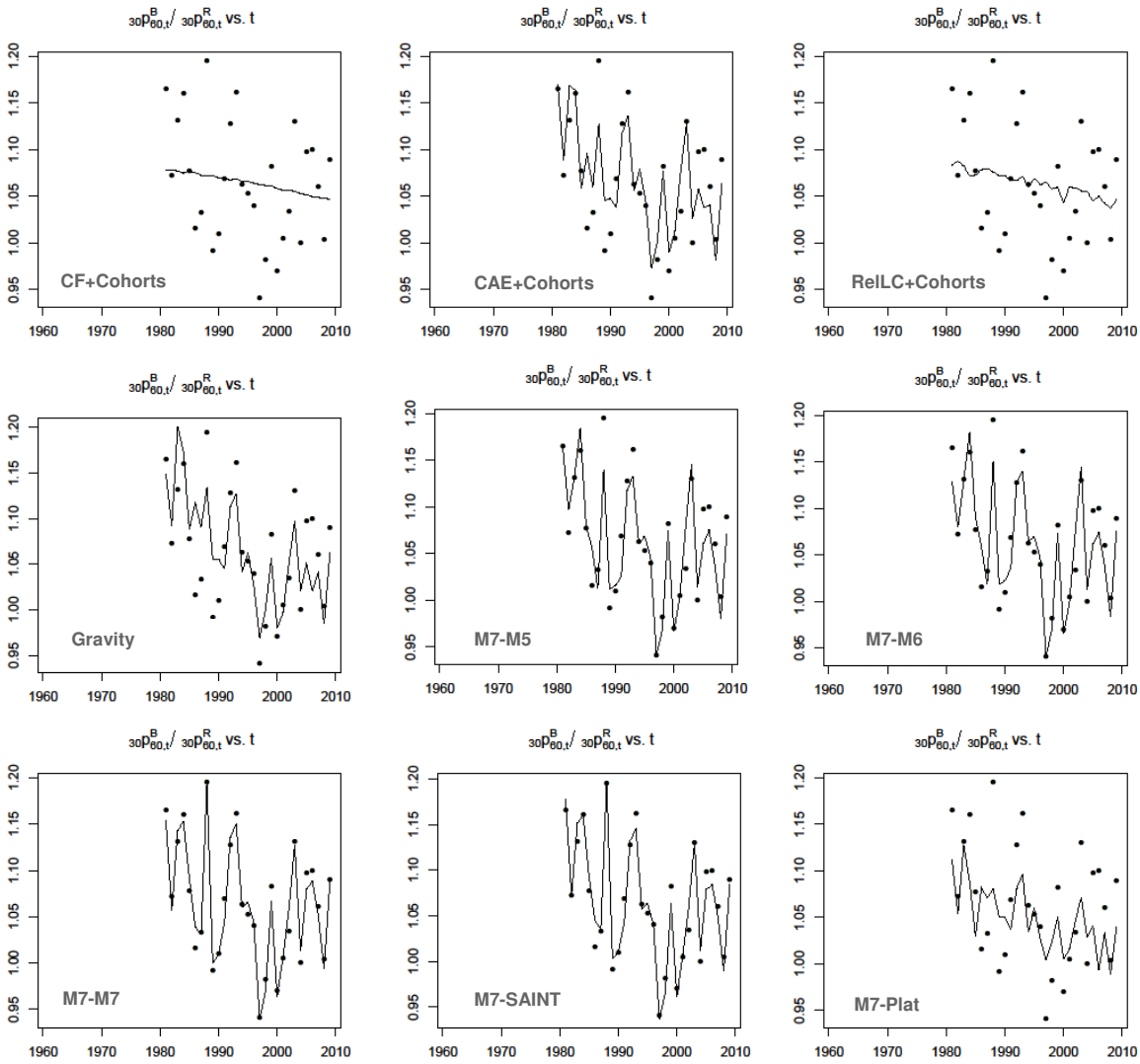


Figure 6.6: Fitted vs. Observed ratio of 30 year period survival probabilities at age 60 for the 'Typical Lives' scheme

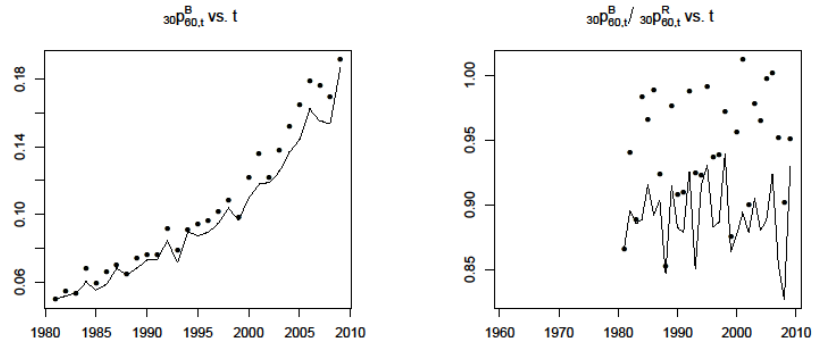


Figure 6.7: Fitted vs. Observed 30 year period survival probabilities at age 60 for the 'Extreme Deprived' scheme using the M7-Plat model

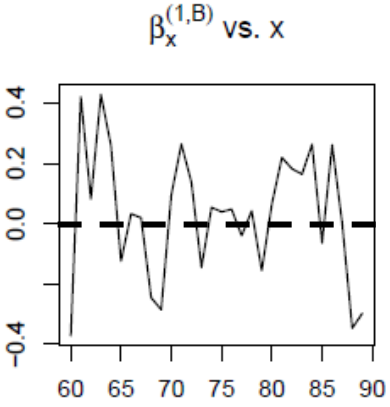


Figure 6.8 Fitted age modulating parameter β_x^B for the RelLC+Cohorts fitted to 'Typical Lives' scheme

6.2.2.5 Trade-off between parsimony and goodness-of-fit (AIC)

The testing of the goodness-of-fit of the models leaves us with six potential candidate models for basis risk assessment. These models are: CAE+Cohorts, Gravity, M7-M5, M7-M6, M7-M7, and M7-SAINT.

The balance between goodness-of-fit and parsimony of these models is investigated in Table 6.5 where we show the AIC values[34] for the book part of each model when applied to the four sample schemes, together with the corresponding ranking across models (in brackets).

From Table 6.5 we note the following:

- CAE+Cohort and M7-M5 models perform very similarly and show the best compromise between goodness-of-fit and parsimony, consistently ranking in the top two places.
- The Gravity model, M7-M6 and M7-M7, which have a book-specific cohort effect, have the worst trade-off between goodness-of-fit and parsimony.

  This suggests that we should generally reject models with a non-parametric book cohort effect on grounds of parsimony. However, for the moment we shall retain such models for further investigation.
- M7-SAINT and M7-M7, which have a quadratic age term in the book model, have a poor trade-off between goodness-of-fit and parsimony.

  This suggests that when considering models from the CBD-Family it is necessary to allow for differences in "level" of mortality and "gradient" by age, but that an additional parameter for a "curvature" by age is not necessary i.e. it is sufficient to inherit the curvature from the reference population. Thus, we eliminate the M7-M7 and M7-Saint models from our list of candidate models.

[34] Recall, the AIC value is computed as $AIC = 2\nu_B - 2\ell_B$ where ℓ_B is the binomial log-likelihood of the book part of the model under the assumption that the reference population is treated as a known offset and ν_B is the number of book-specific parameters of the models

Model	Number of book parameters	Typical Lives	Typical Amounts	Extreme Wealthy	Extreme Deprived
CAE+Cohorts	58	609,078 (2)	593,795 (1)	556,617 (2)	659,538 (1)
Gravity (APC)	114	609,136 (5)	593,866 (5)	556,639 (5)	659,573 (4)
M7-M5	58	609,077 (1)	593,800 (2)	556,604 (1)	659,543 (2)
M7-M6	114	609,136 (4)	593,865 (4)	556,623 (3)	659,597 (5)
M7-M7	142	609,170 (6)	593,895 (6)	556,654 (6)	659,623 (6)
SAINT	87	609,114 (3)	593,830 (3)	556,626 (4)	659,561 (3)

Table 6.5: Effective number of parameters and AIC for the book part of different two-population models fitted to the four test books.

6.2.2.6 Reasonable-forecast level of uncertainty

The outcome of a basis risk assessment exercise will be strongly driven by the expected level of uncertainty around the central forecast of the demographic and financial quantities underlying the index-based hedge. More specifically, the effectiveness of an index-based hedge will be determined by the quantified magnitude of the uncertainty pre and post hedge.

So far, we have shortlisted the CAE+Cohorts, Gravity, M7-M5 and M7-M6 models based on their theoretical properties, practicality and goodness-of-fit performance. However, for basis risk purposes it is crucial to check that these models produce reasonable forecast levels of uncertainty for both single and two population metrics. This entails judging whether or not the forecasted patterns of uncertainty are in line with historical variability.

Following Cairns et al (2011b), we assess this property by examining fan charts of the forecasts produced by the models. Figure 6.9 presents fan charts of 30 year period survival probabilities at age 60 for the England and Wales reference population. Figure 6.10 shows equivalent fan charts of 30 year period survival probabilities at age 60 for the “Extreme Wealthy” test book together with fan charts of the difference between the survival probabilities in the book and the reference population. Each fan chart presents 95% prediction intervals and depicts this forecast output from the stochastic mortality models by dividing the density into 2.5% percentiles. In producing the fan charts we have considered the following sources of uncertainty (risk):

- ***Process risk*** (PR) arising from the possible future trajectories of the time series of the period and cohort indices;
- ***Parameter uncertainty*** (PU) arising from the estimation of the parameters of the model (including those of the time series); and
- ***Sampling risk*** (SR) due to the volatility of the actual mortality experience depending on the size of the population.

In the context of the sources of basis risk introduced in section 1.4, **demographic risk** arises from the combination of process risk and parameter uncertainty.

In practice, process risk is taken into account by simulating trajectories of the period and cohort indices[35], parameter uncertainty is allowed for by using a binomial adaptation of the bootstrapping approach proposed by

[35] To model process risk we use a multivariate adaptation of Algorithm 2 in Haberman & Renshaw (2009) without provision for parameter error. We note that Algorithm 2 in Haberman & Renshaw (2009) is itself an adaptation of the prediction interval approach of Cairns et al. (2006).

Koissi et al (2006)[36], and sampling risk is considered by randomly sampling the number of deaths from the binomial distribution $D_{xt}^{B} \sim Bin(E_{xt}^{B}, q_{xt}^{B})$ once parameter uncertainty and process risk have been taken into account[37]. We note that due to the considerable exposure of the England and Wales population (chosen as the reference population), we have deliberately ignored parameter uncertainty and sampling risk in the reference population.

From Figure 6.9 and Figure 6.10 we can conclude that:

- For all the models the central forecast and their levels of uncertainty for single population metrics are reasonable and consistent in the reference and the book. We note however that there are noticeable differences between the models with M7-M5 and M7-M6 producing significantly higher uncertainty (wider fan widths) than the CAE+cohorts model and the APC (Gravity) model. This reflects the existence of more random period effects in M7-M5 and M7-M6 than in the CAE+cohorts and the APC (Gravity) model.
- The levels of uncertainty in the difference in survival probabilities vary considerably across models and are on the low side when only process risk is contemplated. In particular, the unreasonably tight fan widths of the CF+cohorts confirm the issues with models assuming a perfect correlation between the reference and the book populations.
- The consideration of parameter uncertainty has little impact on single population metrics but makes the confidence intervals in the differences start to look reasonable and in line with the historical volatility.
- Once sampling risk is added the levels of uncertainty still look plausible. However, the differences between the book and the reference populations for some models (e.g. M7-M6) may be considered to lead to levels of uncertainty that are too high in the context of the variation observed historically.

Overall, once all the relevant sources of risk have been included, the four shortlisted models (see figure 6.10) produce plausible forecast levels of uncertainty, but with big enough differences between the models for us to acknowledge model risk as an important issue.

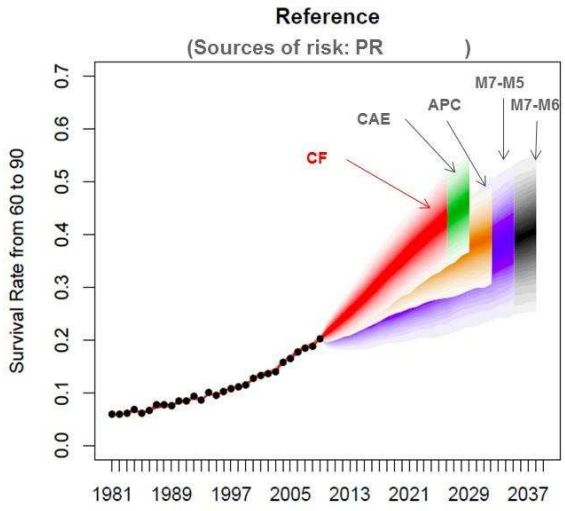


Figure 6.9: Fan charts of 30 year period survival probabilities at age 60 for the England and Wales male reference population using different mortality models. Note that in order to see the fans for different models some are truncated.

[36] We note that in adapting the bootstrap we follow Renshaw & Haberman (2008) and solve for the observed number of deaths instead of the fitted number of deaths as done by Koissi et al (2006)

[37] When taking into account sampling risk we assume that in all future years the book exposure will be equal to the book exposure in the last year of observation.

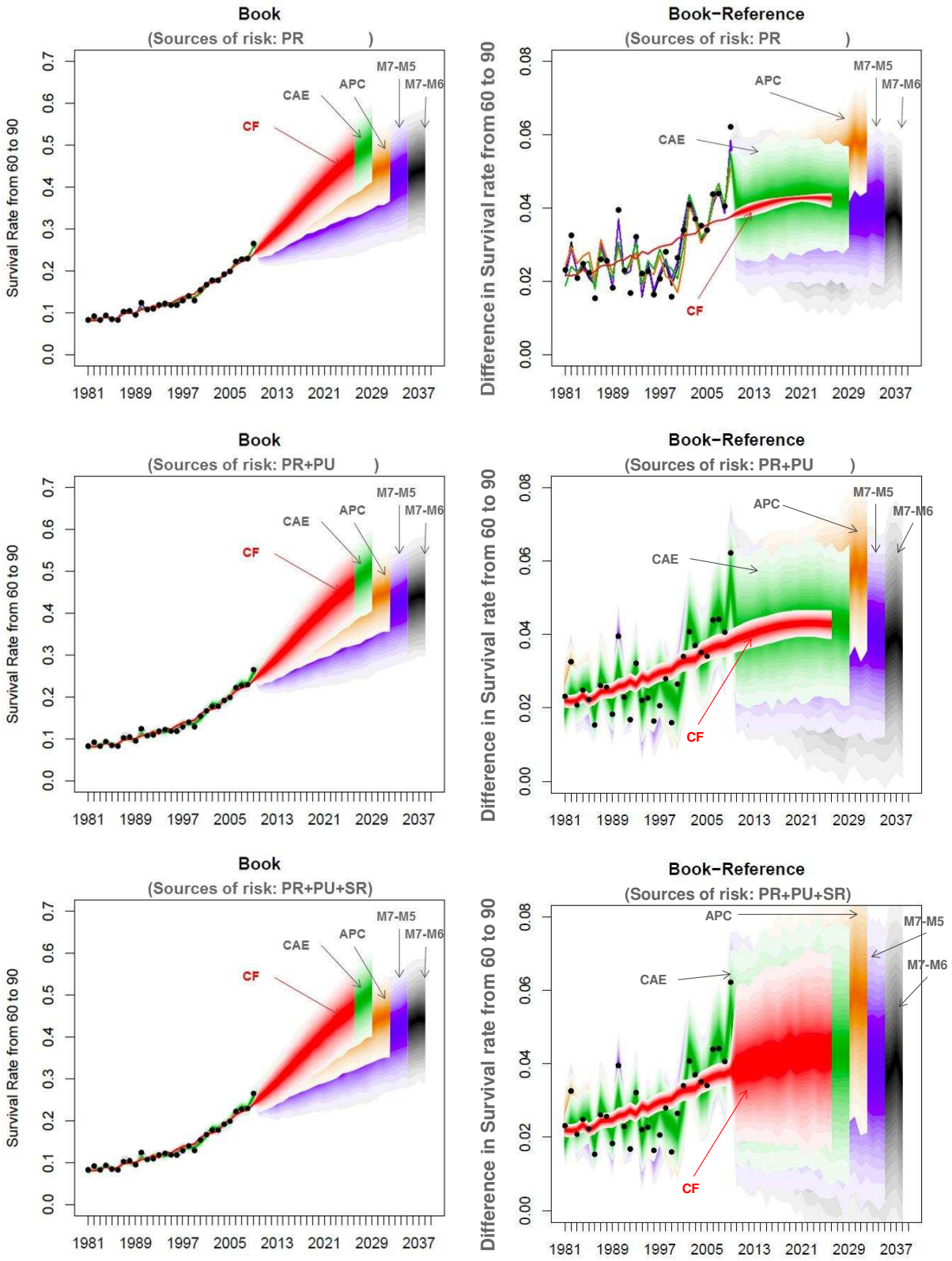


Figure 6.10: Fan charts of 30 year period survival probabilities at age 60 for the “Extreme Wealthy” test book using different mortality models and different sources of risk (PR=process risk; PU=parameter uncertainty; SR=sampling risk). Left panes present results for the book population and right panes results for the difference in survival probabilities in the book and the reference population.

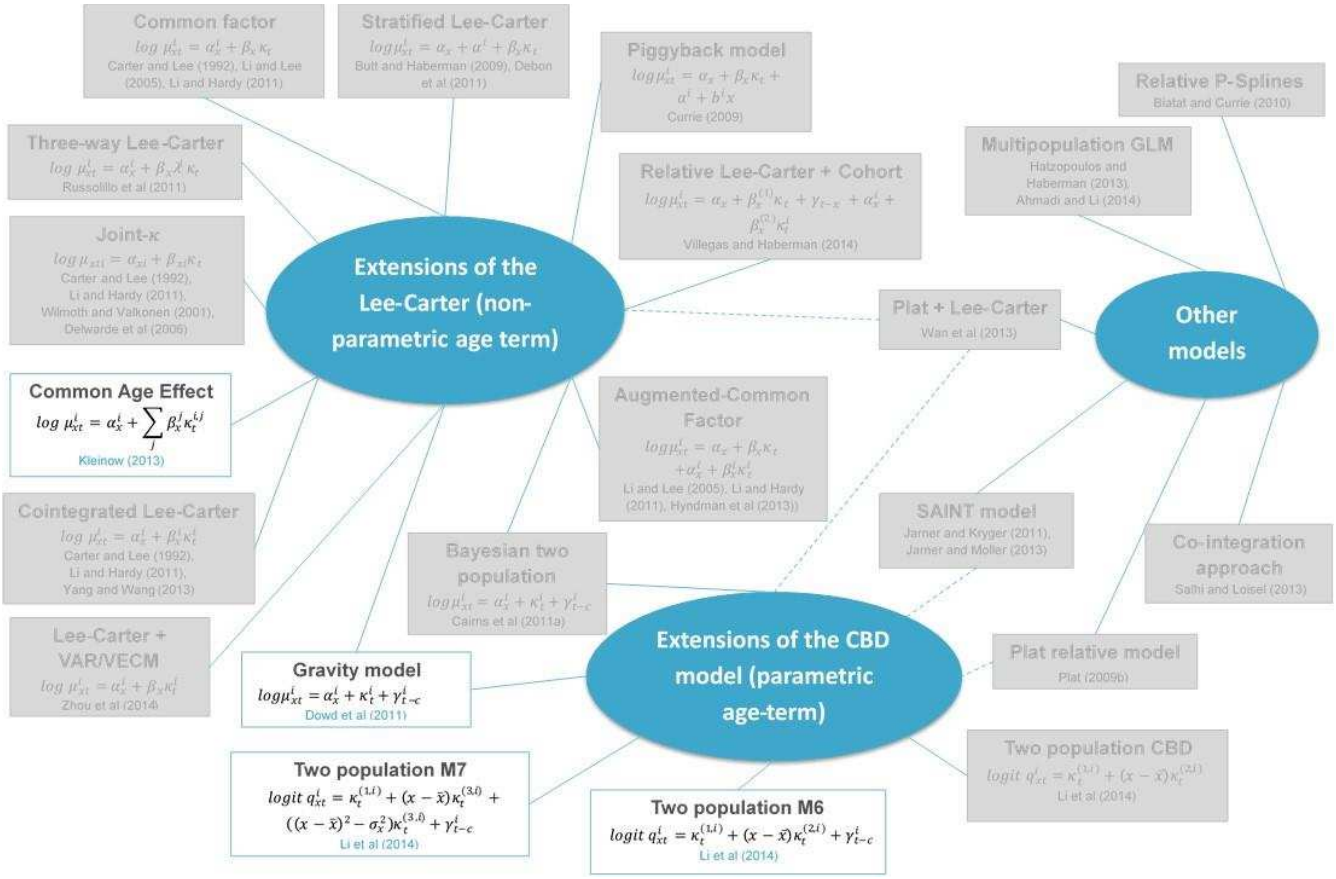


Figure 6.11: Remaining models after stage 2 filtering. Rejected models are greyed-out. Note that in some cases the model tested in this stage 2 differed slightly from the original proposed model (see Section 6.2.2.4)

6.2.3 Stage 3 filtering: Robustness criteria requiring data to assess

In this final stage of assessment we focus on the robustness of the four models which passed the previous levels of filtering to ensure that they perform well in a wide range of circumstances and to test the (lower) limits on data volumes required. In particular, we investigate the robustness of the models with respect to changes in the size of the exposure of the book population, changes in the length of the historical data available, and time-varying socio-economic compositions in the book.

6.2.3.1 Robustness to book size

The smaller the exposure of the population is, the bigger the sampling noise in the data, and the more uncertain the estimates of the parameters of our models are. This additional variability arising from a smaller population size can potentially have a material impact on basis risk assessment.

To explore this phenomenon, we investigate, how the contribution of the different sources of uncertainty to the total level of risk varies by population size. Figure 6.12 decomposes for each of the four shortlisted models:

- the variance of the 30 year period survival probabilities in 10 years' time at age 60 for the "Extreme Wealthy" test book, ${}_{30}p^{B}_{60,2020}$ (middle column); and
- the variance of the corresponding differences with respect to the England and Wales reference population, ${}_{30}p^{B}_{60,2020} - {}_{30}p^{R}_{60,2020}$ (left hand column).

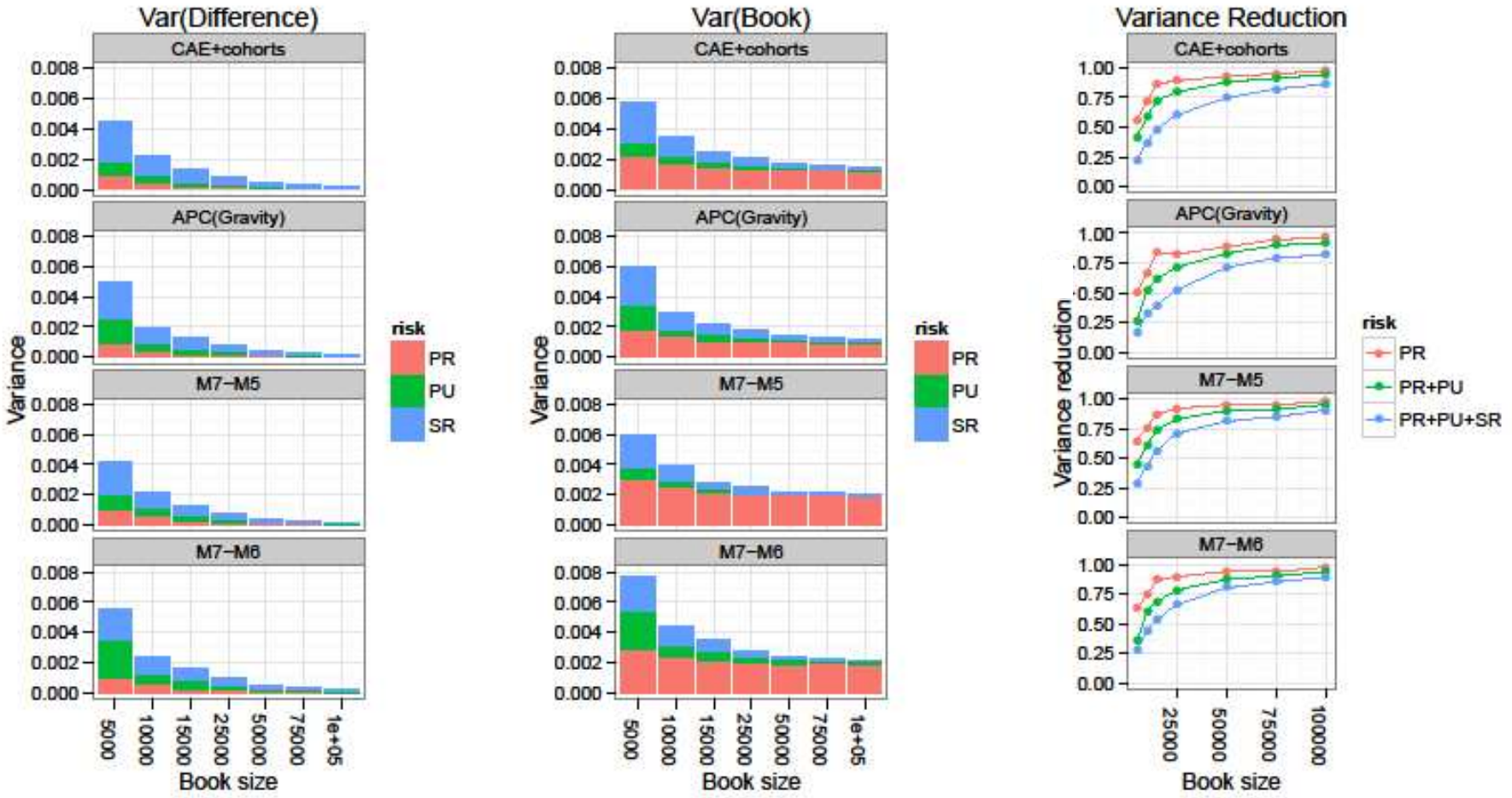


Figure 6.12: Variance decomposition by risk and population size for 30 year period survival probabilities in 10 years' time at age 60 for the "Extreme Wealthy" test book (Left); variance decomposition by risk and population size for the corresponding differences with respect to the England and Wales reference population (Centre); and variance reduction by risk and population size (Right). Book size refers to the total exposure of the book between ages 60 and 90

Figure 6.12 also illustrates for different book sizes the impact of parameter uncertainty and sampling risk on basis risk assessment in terms of the reduction in variance obtained by hedging 30 year period survival probabilities using an index-based swap (right hand column). For each model, combination of risks and book population size, the reported variance reductions are computed as

$$Variance\ reduction = 1 - Var\left({}_{30}p^{B}_{60,2020} - {}_{30}p^{R}_{60,2020}\right)/Var\left({}_{30}p^{B}_{60,2020}\right)$$

We can see how:

- The magnitude of the variance of survival probabilities starts to stabilise around a book size of 25,000 lives. This is particularly noticeable when considering only process risk.
- For book sizes smaller than 25,000 lives, parameter uncertainty is significantly distorting the assessment of basis risk.
- Parameter uncertainty accounts for a significant proportion of the variance for the smaller book sizes.
- For book sizes smaller than 15,000 lives, process risk is unrealistically high distorting the assessment of basis risk and producing artificially low variance reductions
- Models M7-M6 and Gravity which allow for book-specific cohort effects show a significantly higher parameter uncertainty than the CAE+Cohorts model and M7-M5 which do not have a book-specific cohort term.

These observations suggest that, to avoid a distorted assessment of basis risk, the four shortlisted models should only be used when the book exposure is around 25,000 lives. As our analysis here has been based upon men, this book exposure should be considered to apply separately to men and women, unless the book and reference population are to be modelled without regard to gender (see section 9.2.2 for further discussion). In addition, unless there is strong reason to believe in the existence of a different cohort effect in the book to the reference population, the parameter uncertainty in fitting a book-specific cohort term will greatly outweigh any benefits in terms of goodness-of-fit to historical experience.

6.2.3.2 Robustness to history length

In order to assess the impact of changes in history length on the forecasting performance of the models, in this section we carry out a back testing exercise in the spirit of Booth et al. (2006) and Jarner & Kryger (2011, Section 4). This exercise entails the fitting and forecasting of the models using data for the period 1981 to 2010 for different history lengths, book sizes, and IMD compositions in the book population; and the evaluation of different metrics of forecasting performance.

Specifically, the four models were fitted to history lengths ranging from 5 years to 20 years[38], book sizes ranging from 5,000 lives to 100,000 exposed lives between ages 60 to 89 and the four test IMD compositions described before in table 6.1. The performance of the models is evaluated by comparing the actual mortality rates in the book population and actual differences in mortality between the book and the reference population with their corresponding predicted counterparts over the rest of the period until 2010. Forecast bias (actual-fitted) is summarised by averaging across ages, years, book sizes and IMD compositions. The matching absolute errors are also averaged to provide a measure of forecast accuracy.

[38] For instance when considering a history length of 5 years the models were fitted using data for the book population covering the periods 1981-1985, 1983-1987,…, 2003-2007, 2005-2009. In all cases, the reference population data was assumed to start in 1961 and end in the same year as the book population data.

The forecast bias (mean errors) and the forecast accuracy (mean absolute errors) for both rates in the book and differences in rates between the book and the reference, plotted against history length are shown in Figure 6.13. We note the following:

- CAE+cohorts stands out as the best model for forecasting mortality rates in the book with almost no bias on average and with the smallest mean absolute error.
- For differences in mortality rates where we have more than 8 years of for history, the models perform very similarly both in terms of bias and accuracy, with all of them showing a small downward bias.

  History length has a material impact on the out-of-sample performance of the models: For history lengths shorter than 8 years the forecasting performance of the models is poor, particularly for models M7-M5 and M7-M6. The poorer performance of M7-M5 and M7-M6 for the shorter history length is explained by the fact these models have two period indices for the book, implying a more complex and data demanding time series process for the forecasting.

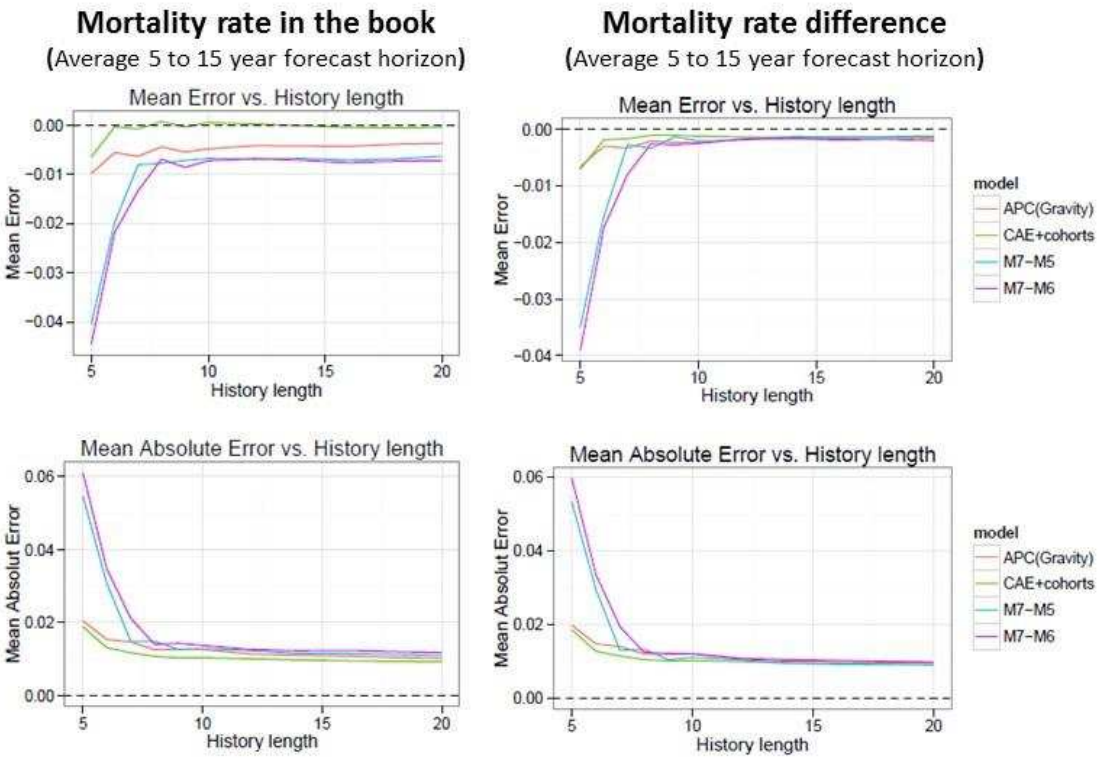


Figure 6.13: Forecasting error in mortality rates in the book (left) and difference in mortality rates between the reference and the book (right). Top panels display mean errors (actual-fit) and bottom panel display results of absolute errors. In all cases, the results are averaged across ages, years, book sizes and socio-economic compositions.

6.2.3.2 Robustness to time-varying socio-economic compositions

So far, in all of our assessments, we have assumed a fixed IMD deprivation split by time in the test books. In order to test the robustness of the models in relation to time varying socio-economic composition in the book, we consider a test book with time varying IMD deprivation split. Specifically, we consider a book exhibiting a *"divergence"* in socio-economic composition with a relatively even split at the start of the period but heavy migration from the two most deprived IMD quintiles towards the two least deprived quintiles (Q1 and Q2) with time (See figure 6.14, next page).

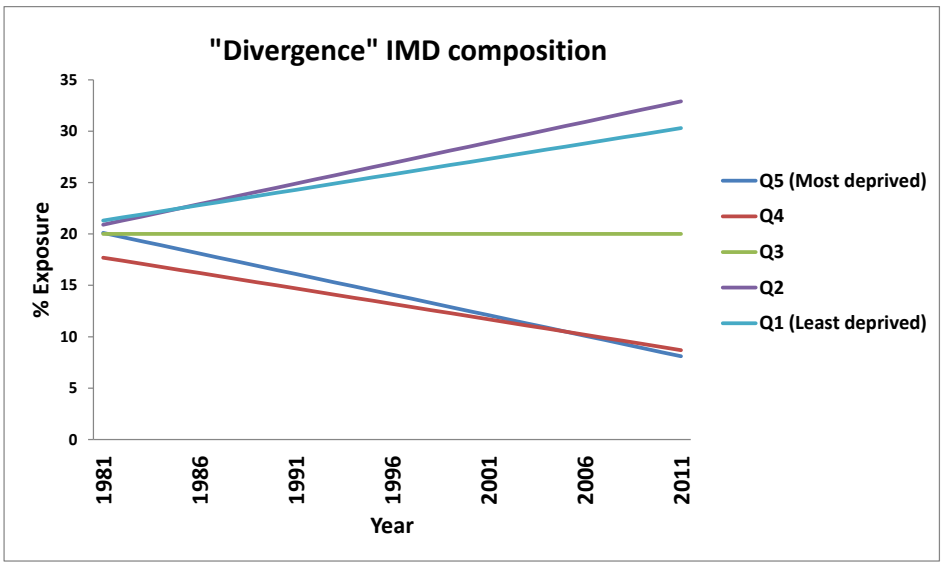


Figure 6.14: Distribution of the exposure by IMD quintile for the book showing a "*Divergence"* in socio-economic composition

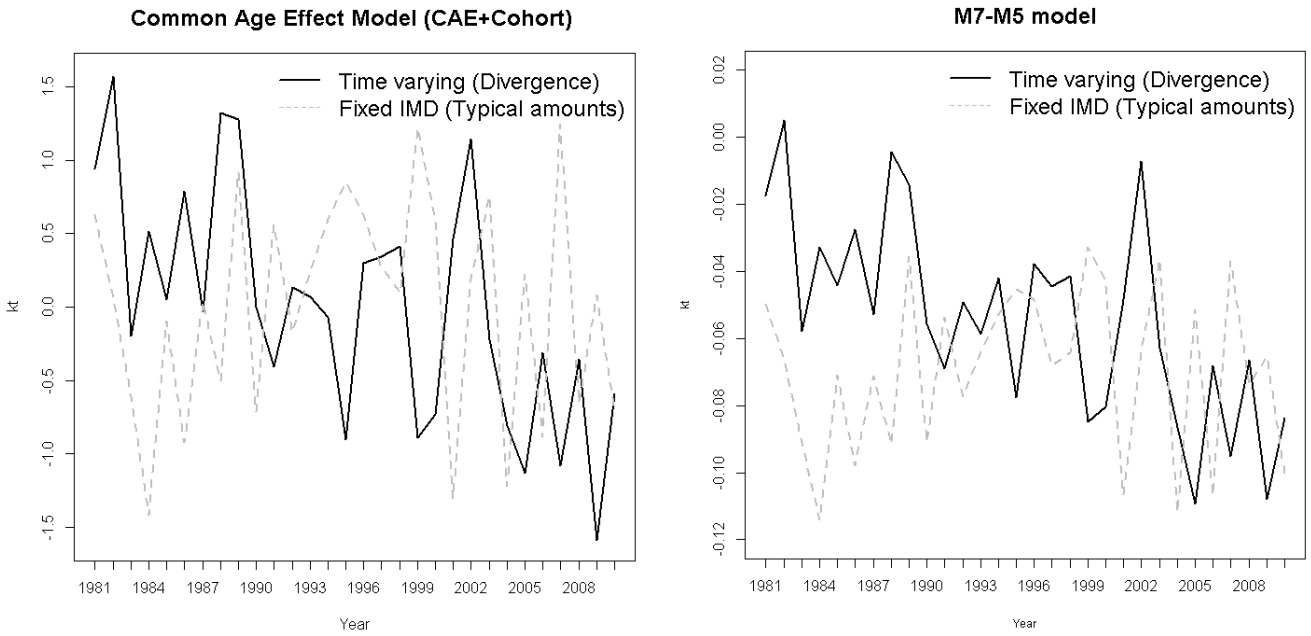


Figure 6.15: First book period index $\kappa_t^{(1,B)}$ for the CAE+cohorts model.and model M7-M5 fitted to test books with a "divergence" IMD pattern and a "Typical amounts" IMD composition

Figure 6.15 (above) plots the historical pattern of $\kappa_t^{(1,B)} \equiv \kappa_t^B$ for the Common Age Effect and $\kappa_t^{(1,B)}$ for the M7-M5 model, for time varying socio-economic mix for the book ("divergence" pattern). For comparison the fixed IMD composition ("Typical amounts") used earlier is plotted as a dotted line.

The book with time varying IMD composition shows a clear downwards trend in the $\kappa_t^{(1,B)}$ terms in contrast to a relatively stable trend for the fixed composition. The downward trend in $\kappa_t^{(1,B)}$ would, in principle, indicate that mortality experience is improving faster in the book than in the reference population. However, this mainly

reflects the changing socio-economic composition of improving average IMD over time so lighter mortality, rather than a true mortality feature in the data. As such, this possible confounding of true improvement differences between the book and the reference populations with changes in the socio-economic mix of the book population introduces challenges when forecasting the book period indices of a two-population model. In such circumstances alternative approaches to direct modelling may be preferable (see section 7).

6.2.4 Summary of assessment

The main conclusions of our systematic assessment of the candidate two-population mortality models for basis risk assessment can be summarised as follows:

- We should not expect any single model to satisfy all possible desirable criteria of a practical solution to assessing basis risk
- However, models M7-M5 and CAE+Cohorts stand out as the models which provide the most suitable balance between flexibility, simplicity, parsimony, goodness-of-fit to data and robustness.
- Both M7-M5 and CAE+Cohorts produce reasonable best estimate projections with plausible levels of uncertainty, but with sufficient differences to recognise model risk as an important issue.
- Unless there is strong reason to believe in the existence of a different cohort effect in the book to the reference population, the parameter uncertainty in fitting a (non-parametric) book-specific cohort term will greatly outweigh any benefits in terms of goodness-of-fit to historical experience.
- The fitting of two-population models should in principle only be pursued when:
    - the book annual exposure is over 25,000 lives – for smaller exposures, the impact of parameter uncertainty may result in a distorted assessment of basis risk
    - there are at least 8 years of reliable book data– for shorter history lengths, the quality of the forecasts is likely to be poor.
- Care needs to be taken when forecasting two-population models fitted to book populations which have undergone significant changes in their socio-economic mix. In these cases, genuine improvement differences may be confounded with changes in the socio-economic mix of the book population.

The above conclusions are underpinned by analysis based on England & Wales population data and the profile of sample schemes drawn from the Club Vita database. We would expect many of the key conclusions to hold for other populations, although specific results (such as AIC rankings) are necessarily dependent on the choice of data. The application of this research to other reference populations and non-UK hedging is considered further in sections 9.1.6 and 9.1.7.

7 A framework for modelling demographic basis risk

In this section we translate our conclusions from the evaluation of the existing two-population models into practical guidelines for the modelling of the demographic basis risk arising from index-based longevity hedges.

We start by proposing a model selection decision tree which provides an easy to use framework for practitioners so that they can identify a suitable approach to assessing the hedge effectiveness of an index-based solution without requiring detailed knowledge of the landscape of multi-population mortality models (Section 7.1).

Next, in Section 7.2, we provide further details on the two mortality models underlying this decision tree. Finally, as we appreciate that in practice some users may have good reason to consider alternative models we discuss some general considerations for the construction of two population models for the assessment of demographic basis risk (Section 7.3).

7.1 Model selection decision tree

The decision tree in Figure 7.1 provides a framework by which practitioners can identify an appropriate approach for modelling the demographic risk associated with an index-based transaction.

By answering a few key questions, users can navigate their way along the tree in order to select the modelling approach which is likely to best suit their particular circumstances (i.e. the demographic and other characteristics of their particular book population). Alternative decision pathways lead to single models which we believe provide the best balance (amongst the wide range of possible models) between flexibility, simplicity, goodness-of-fit to data and robustness to the range of book populations to which it may be applied to.

7.1.1 Two general cases

We have identified two pathways through the decision tree which we believe most practitioners will end up following.

7.1.1.1 A data rich portfolio

The first of these is where the user has a book which is sufficiently large, with a long and stable enough history of data to enable a statistical model to be fitted directly to the practitioners own data. In such circumstances, the user will generally follow the top line of our decision tree resulting in an approach where the M7-M5 model is used to model the reference population and the book.

7.1.1.2 Data does not support direct modelling of the portfolio

We expect that many users will not have sufficient data to follow the direct modelling route. This may be due to insufficient book size (number of lives in the book), lack of historical data, or because the socio-economic mix of the book has changed dramatically over time and so the trends in mortality within the book population are as much attributable to ‘socio-economic drift’ as demographic risk.

In such circumstances, users are likely to apply a “characterisation approach” whereby the book is modelled indirectly by reference to a characterising population with a more reliable and longer mortality experience. Such a characterisation approach is discussed in section 10.

7.1.1.3 Other scenarios

We expect the other scenarios in the decision tree to be less common; although, we have identified an appropriate method to use in each case, and more information is provided on these in section 7.2.

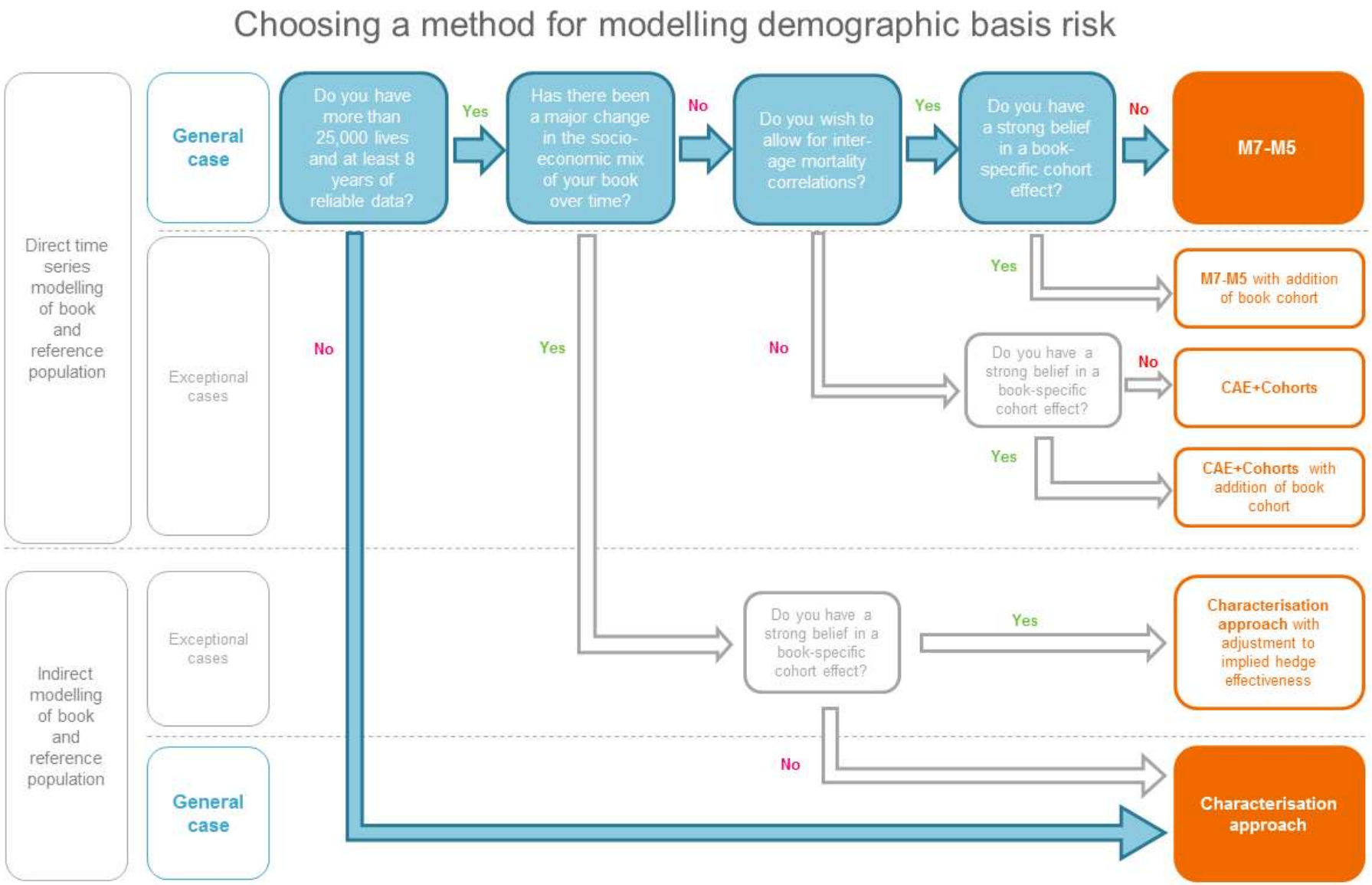


Figure 7.1: Methodology decision tree

7.1.2 The decision tree questions

In the decision tree the user needs to answer four questions (1-4). Taking each in turn:

Question 1: Does the scheme have more than 25,000 lives and at least 8 years of reliable data?

For books with more than 25,000 lives the methods are reasonably robust; both in terms of goodness-of-fit to historical data and forecasting performance. Where the user intends to model the men and women in their book separately (e.g. as using separate indices for men and women) then the 25,000 lives guideline would apply to each gender separately.

The 25,000 lives need not be treated as a very 'hard' cut-off. For example books a little smaller than this e.g. 20,000-25,000 then the user is likely to still find direct modelling informative. However, users with smaller book sizes are likely to find material parameter uncertainty when trying to fit models directly, poor goodness-of-fit and forecasting performance. As the book size falls, so users will also find that sampling risk dominates the assessment of basis risk. Here, users should favour a characterisation approach rather than trying to fit models directly to the experience data. In particular for a very small book size (say, less than 10,000 lives) parameter uncertainty and sampling risk will be too high and a characterisation approach will be unavoidable in order to obtain a meaningful measure of basis risk / hedge effectiveness.

Similarly, where the book has less than 8 years of reliable data direct modelling using purely the book data is not practical. The issue here is the limited data with which to fit the coefficients of a time series – indeed, in extremis, there may be no reliable experience data. A characterisation approach – which can leverage the long back history of an alternative data source – is likely to be preferable.

Since different considerations[39] are driving the thresholds in the lives and back-history dimension, any scope to 'offset' more data in one dimension for less in another (e.g. more lives and less back history) is limited.

Question 2: Have there been any major changes in the socio-economic mix in the book over time?

Where a book has undergone very significant shifts in socio-economic mix it is very likely that any fitted time series coefficients will reflect both improvements in mortality for the book's average socio-economic mix and the changing mix over time.

For example, suppose blue-collar workers experience double the mortality rates of white-collar workers, and over the period in question the book's mix changed from 100% blue-collar to 100% white-collar workers. In such circumstances the book would show a 50% improvement in mortality even if there has been no change in the mortality of blue or white collar workers.

In practice, most books will have only undergone modest changes in socio-economic mix over time and so this is unlikely to be an issue. However, where the book has seen a fundamental shift in the underlying population – for example a change in target market for an insurer, or a major change in the nature of business for a pension scheme sponsor – then we suggest using the characterisation approach. This is because direct modelling of the trend in the difference between the book and the reference would usually assume:

(i) Either some reversion to an average socio-economic mix (if, as commonplace, the VAR(1) time series is used); or

(ii) A continuation of the historical trend in shifting socio-economic mix (if a random walk with drift or other times series with a trend is used).

By applying a characterisation approach, the user is able to make more of an explicit assumption as to how the socio-economic mix will change over time[40].

[39] namely levels of parameter uncertainty in both the trends and levels of mortality vs ability to fit times series respectively

Question 3: Do you wish to allow for inter-age mortality correlations?

The evaluations of models in section 6 highlighted M7-M5 and CAE+cohorts model as the best performing models against a range of desirable criteria, with the CAE+cohorts outperforming the M7-M5 model in terms of goodness-of-fit and back-testing performance. However, a key weakness of the CAE+cohorts[41] model is that it imposes a very simple structure to the correlation between mortality rates at different ages compared to models which use more than one period time index such as M7-M5.

In choosing between the two alternative paths at this point in the decision tree, the user should take into account the following considerations:

- **Structuring**: When structuring a longevity hedge, it is likely to be important to have a rich structure to inter-age mortality correlations. In most cases the user will be looking to hedge a book population covering a wide range of ages. Where a model assumes perfect correlation between changes in mortality at different ages it could lead the user to conclude that forward contracts for a single age (such as age 60) provide just as good a hedge as holding contracts at a range of different ages. In such scenarios, the user should favour the richness of the M7-M5 model.
- **User's own view of existing longevity risk?** In many instances the user will want to assess the indicative reduction in basis risk following the implementation of an index-based hedge, without necessarily considering in detail the precise structuring of the instruments held. In such cases it will suffice to compare the uncertainty pre- and post-hedge.

  Where the user does not have an existing method for assessing the longevity risk in the book population, then using a modelling framework which provides a reliable assessment of longevity risk for both the book population and the difference between the book and reference population (i.e. post hedge exposure) will be important. In such circumstances, the user may prefer the CAE+cohorts model which we have found to perform better for forecasting the individual populations.
- **Short history length**: Where the history length is on the short side, users may prefer the CAE+cohorts model as its forecasting is less data demanding.
- **Model risk:** Some users may want to consider both paths of the decision tree at this stage as a way of assessing model risk.

A detailed comparison of the main features of models M7-M5 and CAE+cohorts is presented in section 7.2.3.

Question 4: Do you have a strong belief in a book-specific cohort effect?

In general a (non-parametric) book-specific cohort effect cannot be justified on grounds of parsimony (balance between simplicity and goodness-of-fit to historical data). However, some users may have good reason to believe that there is a material book-specific cohort effect different to that inherited from the reference population. In such circumstances, the user is recommended to adapt the model best suited to his/her requirements for the richness of the inter-age correlations by including a book-specific cohort effect. We would suggest that (generally) such a cohort effect should have a parametric form as it is unlikely that users' data would support the fitting of a non-parametric form without considerable parameter uncertainty.

[40] An exception to this guidance would be where the user believes this shift will continue to apply in the future to the particular population covered by the hedging instrument – however as such populations tend to be a specific cohort of lives the potential for ***continued*** major shifts in the socio-economic mix would generally be far less.

[41] Note that a richer correlation structure can be obtained if the CAE+cohorts model is further extended with the addition of an extra bilinear term. However this raises additional challenges in terms of estimating the parameters and applying identifiability criteria.

7.2 Direct modelling – choosing the model

The decision tree proposes the M7-M5 and the CAE+cohorts models as the default statistical models for the direct modelling branches of the tree. In this section, we describe in detail these two models and compare their features.

7.2.1 Parametric form for shape of mortality with age ('M7-M5')

The M7-M5 model is a two-population extension of the Cairns-Blake-Dowd (CBD) model of mortality introduced in Cairns, Blake, & Dowd (2006). The single population CBD model can be expressed as:

$$\text{logit } q_{xt} = \kappa_t^1 + (x - \bar{x})\kappa_t^2 \qquad (7.1)$$

where $\bar{x}$ is the average age in the data and κ_t^1 and κ_t^2 are random period effects. The CBD model, also known as M5, assumes that for a fixed calendar year t the logit death rates are linear functions of age with the level at $\bar{x}$ determined by κ_t^1 and slope κ_t^2. In contrast to the Lee-Carter model (which considers only one period factor), the CBD model includes two period factors. As a result, the model can capture the imperfect correlation structure in mortality rates improvements across ages.

Model M7, introduced by Cairns et al. (2009), is an extension to the original CBD model in which a quadratic age term and a cohort effect are added. Mathematically, M7 is given by,

$$\text{logit } q_{xt} = \kappa_t^1 + (x - \bar{x})\kappa_t^2 + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^3 + \gamma_{t-x} \qquad (7.2)$$

where σ_x^2 is the average value of $(x - \bar{x})^2$.

In the context of two population mortality modelling CBD type models have been used by Li et al (2014). For instance, they consider two populations where an M7 model is fitted independently to each of the two populations:

$$\text{logit } q_{xt}^i = \kappa_t^{(1,i)} + (x - \bar{x})\kappa_t^{(2,i)} + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^{(3,i)} + \gamma_{t-x}^i, \quad i = 1,2 \qquad (7.3)$$

and the relationship between the populations is considered through the joint modelling of the period indices $\kappa_t^{(j,i)}, j = 1,2,3$.

Starting from the two population M7 in equation (7.3), we have made the following considerations to obtain the M7-M5 under the general relative formulation described in section 5.2:

- Assume that the curvature term and the cohort effect are common for the two populations as in our model testing it was found that these differences were not statistically significant for most populations.
- Assume, thus, that

$$\text{logit } q_{xt}^R = \kappa_t^{(1,R)} + (x - \bar{x})\kappa_t^{(2,R)} + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R \qquad (7.4)$$

  and,

$$\text{logit } q_{xt}^B - \text{logit } q_{xt}^R = \kappa_t^{(1,B)} + (x - \bar{x})\kappa_t^{(2,B)} \qquad (7.5)$$

  which is equivalent to

$$\text{logit } q_{xt}^B = \left(\kappa_t^{(1,R)} + \kappa_t^{(1,B)}\right) + (x - \bar{x})\left(\kappa_t^{(2,R)} + \kappa_t^{(2,B)}\right) + ((x - \bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^R \qquad (7.6)$$

In (7.4) $\kappa_t^{(1,R)}$, $\kappa_t^{(2,R)}$, $\kappa_t^{(3,R)}$ are stochastic period effects driving mortality change in the reference population and γ_{t-x}^R captures cohort effects. From (7.6), it is clear that $\kappa_t^{(1,B)}$ captures differences over time between the

reference and the book populations in the general level of mortality, while $\kappa_t^{(2,B)}$ captures differences of the age-slope of mortality in the book population relative to the age-slope of mortality in the reference population.

The parameters of Model M7 are not uniquely identified as there are a variety of equivalent solutions to equation (7.4). For example we can switch from γ_{t-x}^R to $\tilde{\gamma}_{t-x}^R = \gamma_{t-x}^R + \phi_1 + \phi_2(t-x-\bar{x}) + \phi_3(t-x-\bar{x})^2$ and corresponding adjustments to $\kappa_t^{(1,R)}$, $\kappa_t^{(2,R)}$, and $\kappa_t^{(3,R)}$, with no impact on the fit of the q_{xt}^R.

In order to be able to fit the model we therefore need some identifiability constraints. We use the standard constraints as per Cairns et al (2009):

$$\sum_c \gamma_c^R = 0$$

$$\sum_c c\gamma_c^R = 0$$

$$\sum_c c^2\gamma_c^R = 0$$

where $c = t - x$. The consequence of these constraints is that γ_c^R will fluctuate around 0 and will have no discernible linear trend or quadratic curvature.

7.2.2 Non-parametric form for shape of mortality with age ('CAE+cohorts')

A wealth of multi-population extensions of the Lee-Carter model have been proposed recently in the mortality modelling literature. For instance, Kleinow (2013) considered the Common Age Effect (CAE) model where the mortality of population i, $i = 1, \dots, K$, is modelled as[42]

$$\log\mu_{xt}^i = \alpha_x^i + \sum_j \beta_x^j \kappa_t^{i,j}$$

The main feature of this model is that there is a common set of age-response parameters (β_x^j) across the different populations. Here, α_x^i represent the age-specific mortality pattern of population i and the $\kappa_t^{i,j}$, $j = 1, \dots, J$ are stochastic period effects driving mortality change in population i.

In order to adapt the original formulation of the CAE and derive the CAE+cohorts model under the general relative formulation described in section 5.2 we have made the following considerations:

- Reformulate the model in terms of logit of mortality probabilities (q values).
- Consider a single bi-linear term and add a common cohort effect in the reference and book populations.
- Assume, thus, that

$$\text{logit}\, q_{xt}^R = \alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R \qquad (7.7)$$

and

$$\text{logit}\, q_{xt}^B - \text{logit}\, q_{xt}^R = \alpha_x^B + \beta_x^R \kappa_t^B \qquad (7.8)$$

In (7.7) α_x^R captures the general age-specific mortality pattern in the reference population, κ_t^R is a stochastic period effect driving mortality change in the reference population, β_x^R measures the age-specific response to changes in κ_t^R and γ_{t-x}^R captures cohort effects.

[42] A similar model was considered by Zhou, Li, & Tan (2013).

Equation (7.8) is equivalent to

$$\text{logit}\, q_{xt}^{B} = (\alpha_x^R + \alpha_x^B) + \beta_x^R(\kappa_t^R + \kappa_t^B) + \gamma_{t-x}^R \quad (7.9)$$

which implies that: α_x^B determines the average level mortality difference between the book and the reference populations, and κ_t^B determines the deviations of the mortality improvements of the book population around the mortality improvements of the reference population.

Similarly to model M7-M5, the CAE+cohorts model has an identifiability problem. For example multiplying each β_x^R by a constant b and each κ_t^R by $\frac{1}{b}$ will yield the same fit. To ensure identifiablity we use the following constraints:

$$\sum_x \beta_x^R = 1 \quad (7.10)$$

$$\sum_t \kappa_t^R = 0 \quad (7.11)$$

$$\sum_c \gamma_c^R = 0 \quad (7.12)$$

$$\sum_t \kappa_t^B = 0 \quad (7.13)$$

These constraints imply that α_x^R can be interpreted as the average level of mortality in the reference population across the period of the reference data and that α_x^B can be interpreted as the average mortality deviation of the book from the reference across the period of the book data.

In order to improve the stability and robustness of the CAE+cohorts used for the reference population, we also add the constraint

$$\sum_c c\gamma_c^R = 0$$

as suggested in Hunt & Villegas (2014). This ensures that α_x^R adheres to the typical shape of a life table, and that γ_c^R will fluctuate around 0 with no discernible linear trend. In addition, when κ_t^R is well approximated by a straight line as is the case of the England and Wales experience, this constraint will have minimal impact on the fit to data obtained by the model.

7.2.3 Comparing the two models

In choosing between the M7-M5 and the CAE+Cohorts, it is worth taking the following considerations into account.

7.2.3.1 Ease of use

Both models require the user to fit to historical data and then forecast.

- The M7-M5 model is easy to fit to historical data as it can be formulated under a GLM framework. In particular, the M7 reference part is a binomial GLM model which is estimated via maximum likelihood using standard statistical software (see for instance Currie (2014)). In addition, conditional on the reference population parameters, the M5 part is also a binomial GLM and can be estimated similarly. By contrast, the estimation of the reference part of the CAE+cohorts model is less straightforward because of the bilinear term $\beta_x^R\kappa_t^R$ and the known robustness and stability issues of cohort extensions of the Lee-Carter model (see e.g. Cairns et al. (2009); Hunt & Villegas (2014)).

- The CAE+cohorts model uses univariate time series for both the period effects within the reference population and the book population, compared to the multivariate time series used by the M7-M5 model. This means that the CAE+cohorts is easier to use and potentially requires a shorter back history when it comes to generating the time-series for forecasting.

Overall, we see little difference between the two models for ease of use and both could readily be programmed in a freeware format (akin to the LifeMetrics Excel add-in[43]).

7.2.3.2 Richness of correlation structure

Within the CAE+cohorts model, there is a simple correlation structure between annual changes in mortality at different ages. In particular the existence of a single period effect for the reference population implies that there is perfect correlation at all ages except at the youngest ages, where there is potentially additional randomness arising from the arrival of new cohorts with an unknown cohort effect (see Cairns et al (2009)). In contrast, M7-M5 model allows for imperfect correlation between annual changes in mortality at different ages due to the presence of multiple period factors in the reference population.

7.2.3.3 Applying known base rates

Both of the models produce base rates for the mortality of the book population (i.e. the mortality rates by age for a recent point in time). In practice though, the user may have an alternative approach to the base rates which he or she would prefer to use – for example due to an established mechanism for assessing this such as experience analyses on the book, or due to using a more granular method.

Given the natural interpretation of α_x^B as mortality level difference under the CAE+cohorts model, the superposition of the user's preferred base mortality rates can be easily done (see section 9.2.1.1).

By contrast, the M7-M5 model does not directly allow the superposition of the user's preferred base rates. Nevertheless, this can be achieved by either:

- including an age specific non-parametric term as done in Plat (2009a), which may, however, complicate considerably the identifiability of the model (see Hunt & Blake (2014));
- applying the year on year implied improvements at each age produced by the book projection to the user's known base rates (see section 9.2.1.2)

7.2.3.4 Extension to older ages

In practice the user is likely to want to be able to model mortality, and basis risk, across the age spectrum, including older ages beyond those to which there is adequate data to reliably fit the models. In the case of the M7-M5 model there is a 'natural' extension in so far as the functional form could be applied at older ages – however care would be needed as the curvature term could create mortality tables which might be deemed as biologically unreasonable (i.e. mortality declining with age). In the case of the CAE+Cohorts model more thought would be needed as there is no 'natural' functional extension. In practice though, some form of subjective (structural) assumption is likely to be needed under both models.

7.3 General comments on constructing a two-population model

Our previous sections have suggested the use of either M7-M5 or the CAE+cohorts when undertaking a direct modelling exercise of the mortality of the reference and the book populations. However, this need not preclude a user from considering additional models. Indeed users may wish to look at alternative models as part of sensitivity testing; or in order to gain a better understanding of model risk; or to err on the side of adding more features into the model than historic back-testing alone might suggest are needed as part of a personal belief on

[43] Available at http://www.macs.hw.ac.uk/~andrewc/lifemetrics/

the 'complexity' of mortality. Further as time goes on new models will enter the actuarial literature – and our work to date can help users in integrating those models into a basis risk assessment.

Therefore, in this section we provide some general guidelines for the construction of alternative two-population models for basis risk assessment.

When building a two population model for assessing longevity basis risk, it is usual to find that the reference population is considerably larger and has a longer back history of data than the book population. It is therefore natural to start by selecting an appropriate model for the reference population since:

- the larger reference population will tend to influence the book mortality but not the other way round
- good models for the reference population will generally already exist and so will be a useful source of information when modelling the book population

Once the reference population model is chosen a reasonable approach would be to select the book part of the model from within the same model family of the reference part. This is because:

- modelling the reference parts provides useful insights and model structures when constructing models for the book; and
- there is a correspondence between the model parameters in the book and the reference populations which makes interpretation of the parameters consistent and makes the subsequent analysis more comprehensive and coherent in both populations.

Our research on different models has also identified that:

- it is desirable to include at most two book-specific time-dependent terms
- any parameter which moderates the sensitivity of the book to these time trends at different ages should be inherited from the reference book (i.e. $\beta_x^{(i,B)} \equiv \beta_x^{(i,R)}$)[44]
- it is generally appropriate not to include a book specific cohort effect

In mathematical terms, if the preferred reference population model is given by

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

then a good starting point for the book model would in general be of the form:

$$logit(q_{xt}^B) - logit(q_{xt}^R) = \alpha_x^B + \sum_{i=1}^{M} \beta_x^{(i,R)} \kappa_t^{(i,B)}$$

We would usually expect M to be at most 2 as it is unlikely that the book population can support more than two time series i.e. $M \leq \min(2, N)$[45].

By way of example, if the user chooses to model the reference population using[46]

[44] As noted previously in section 6.2.2.4 and illustrated in Figure 6.8 the paucity of book data can hardly support the estimation of β_x^B without resulting in non-robust and erratic parameter estimates.

[45] Note that the M7-M5 model and the CAE+cohorts can be derived from this form by applying the previous rules if we start by modelling the reference population using a M7 model or a Lee-Carter +Cohorts model, respectively.

[46] As described in Börger et al (2013)

$$logit(q_{xt}^R) = \alpha_x^R + \kappa_t^{(1,R)} + (x - \bar{x})\kappa_t^{(2,R)} + \left(x_{young} - x\right)\kappa_t^{(3,R)} + (x - x_{old})\kappa_t^{(4,R)} + \gamma_{t-x}^R,$$

where x_{young} and x_{old} are predefined constants, then the suggested book model would be

$$logit(q_{xt}^B) - logit(q_{xt}^R) = \alpha_x^B + \kappa_t^{(1,B)} + (x - \bar{x})\kappa_t^{(2,B)}$$

8 Case study – Direct modelling

8.1 Introduction

This case study illustrates how practitioners can use the decision tree in figure 7.1 to select a modelling approach which may be a good starting point in their particular circumstances. We then illustrate the methodology using a simple measure for the level of longevity risk before and after an index-based hedge.

The case study is based on men from a sample scheme drawn from the Club Vita dataset (See Appendix A) which has the following characteristics:

- Contains between 20,000 and 27,000 male lives for each year under consideration
- Has relatively long back history, with good quality data covering the period 1993-2011
- Consists of a lower socio-economic composition than the reference population (England & Wales)

8.2 Model selection process

Question 1: Does the scheme have more than 25,000 lives and at least 8 years of reliable data? - YES

The chosen scheme has between 20,000 and 27,000 lives available each year and a history length of 18 years so is a good candidate to be modelled directly.

Question 2: Have there been any major changes in the socio-economic mix in the book over time? - NO

In general, trustees and insurers have a deep understanding of the nature of their schemes and should therefore be in a good position to assess whether the socio-economic mix has substantially changed over time. In this case, there have been no major changes in the nature of the business of the sponsor of the pension scheme.

Figure 8.1 confirms that while there has been a small drift away from the most deprived IMD, there has been no fundamental change. (The figure also highlights how this scheme has a bias towards areas of higher deprivation.)

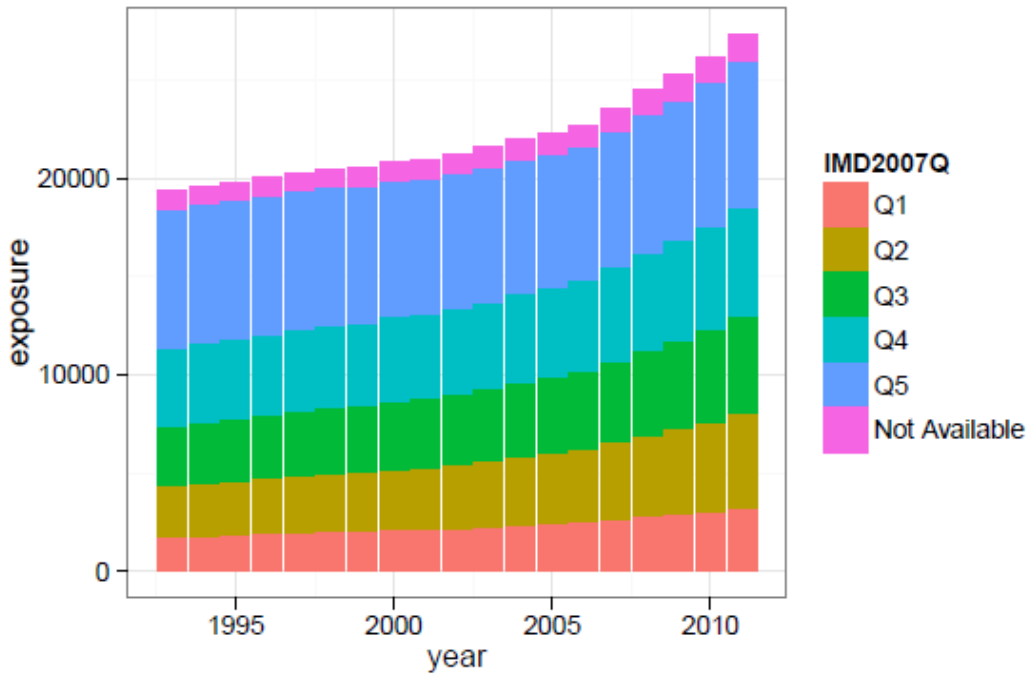


Figure 8.1: Development of the socio-economic mix over time in the sample scheme used in this case study.

Question 3: Do you wish to allow for inter-age mortality correlation structure? - YES

In our case study the scheme is looking to structure an index-based swap and so requires an inter-age correlation structure.

Question 4: Is there a strong belief in a book-specific cohort effect? - NO

The scheme has no reason to believe that there is a material book-specific cohort effect different to that inherited from the reference population. Therefore there is no need to allow for any such an effect in this case study modelling.

Conclusion

By following the decision tree in figure 7.1, our answers to the four questions lead us to fit the data from the sample scheme to the M7-M5 model.

8.3 Model fitting

In this section we illustrate the results of fitting the M7-M5 model to the historical data for the book and reference populations.

8.3.1 M7-M5 Model – Fitted parameters

Reference: $logit(q^R_{xt})$	Book: $logit(q^B_{xt}) - logit(q^R_{xt})$
$\kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma^R_{t-x}$	$\kappa_t^{(1,B)} + (x-\bar{x})\kappa_t^{(2,B)}$

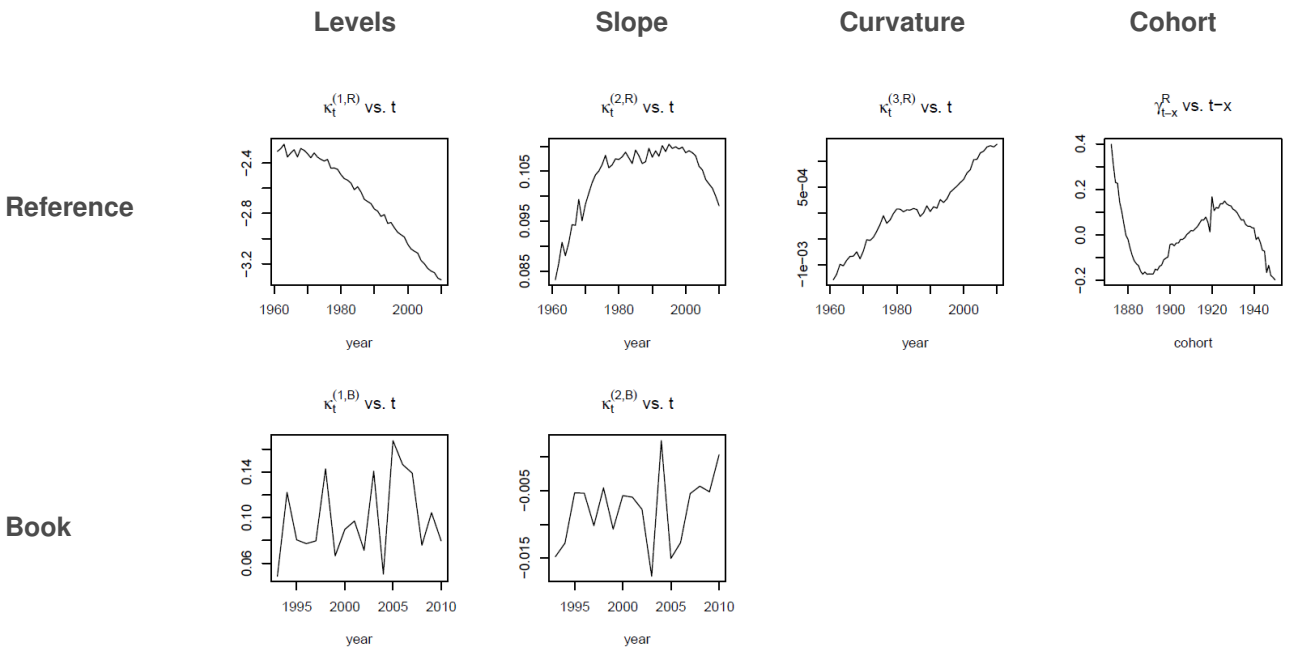


Figure 8.2: Fitted parameters of the M7-M5 model using the sample pension scheme.

Figure 8.2 shows the parameters of the M7-M5 model when fitted to the sample scheme from Club Vita and the reference population (England & Wales). The top row shows the fitted parameters for the M7 model fitted to the reference population and the bottom row represents the parameters for M5 model fitted to the difference between the book and the reference population.

Starting with the reference population, we can see how:

- the $\kappa_t^{(1,R)}$ parameter shows a clear downward sloping trend which demonstrates steady mortality improvements in the reference population over time
- $\kappa_t^{(2,R)}$ reflects how the age effect of mortality changes with time
- $\kappa_t^{(3,R)}$ shows how the curvature of mortality rates with age changes with time
- γ_{t-x}^{R} reflects any cohort related effects in the reference data. We can see the faster decline in mortality for birth cohorts between 1925 and 1945 which coincides with the well-known UK golden generation

The book parameters define how the book mortality differs from that of the reference population. The positive values of $\kappa_t^{(1,B)}$ reflect the higher mortality rates in the book than in England & Wales (consistent with the weight towards the more deprived areas). The negative values of $\kappa_t^{(2,B)}$ moderate these higher mortality rates by reducing the slope of the resulting mortality curves over age, helping to maintain a compensation law[47] of mortality whereby the mortality of the book and reference populations converges at older ages.

8.3.2 M7-M5 Model – Time varying trend

When modelling future rates using the M7-M5 model in section 5.2.3 a multivariate random walk with drift (MRWD) was used for the reference population and, given the absence of any clear trends over time for $\kappa_t^{(1,B)}$ and $\kappa_t^{(2,B)}$ in figure 8.2, a vector-autoregressive process of order 1 was used for the book population i.e.:

$$\boldsymbol{\kappa}_t^{\mathrm{B}} = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1}\boldsymbol{\kappa}_{t-1}^{\mathrm{B}} + \boldsymbol{\xi}_t^{\mathrm{B}}, \qquad \boldsymbol{\xi}_t^{\mathrm{B}} \sim N(\mathbf{0}, \Sigma^B)$$

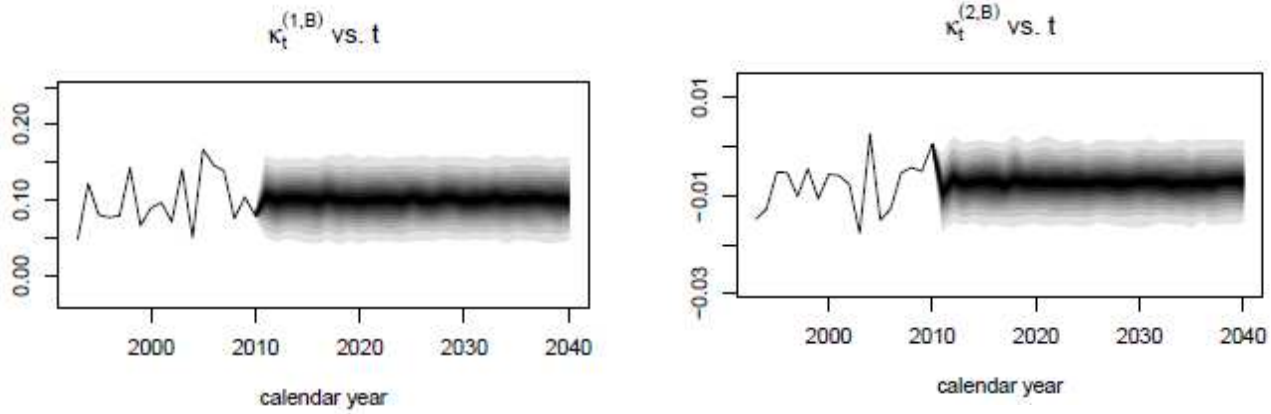


Figure 8.3: Projected time indices of the M7-M5 model fitted to a sample pension scheme with VAR(1) process

Figure 8.3 demonstrates the behaviour of a full simulation of the $\kappa_t^{(1,B)}$ term when projected over 30 year time horizon. By using an autoregressive model of order 1 to model future mortality rates, we assume that $\kappa_t^{(1,B)}$ reverts to its historical mean. This seems a reasonable assumption with respect to the trend in the fitted historical rates and is consistent with typical practice; however, this assumption and alternatives are considered further in Section 9.

[47] See Gavrilov & Gavrilova (1991)

8.4 Hedge effectiveness results

Having determined a suitable parameterisation of the M7-M5 model the book and reference population are simulated to generate 10,001 model points. The simulations allow for process, parameter and sampling risk as described in section 6.2.2. We illustrate the output of the model in the form of a simple hedge effectiveness metric, specifically the variance reduction in 30 year survival probabilities at age 60 with 10 year time horizon[48]. The graphics to the right show the impact of an index-based swap on a book of pensioners currently in payment.

Figure 8.4 shows simulated density plots where the x-axis represents the range of outcomes for the 30 year survival probabilities relative to the mean and the y-axis shows the density. The black line represents the unhedged book (i.e. the book population as simulated). The pink line shows the difference between the book and reference population and so illustrates the case when an index-based hedge is applied to the book. In each case an allowance is made for all three sources of uncertainties in the model (process risk, parameter risk and sampling risk). The tails for the hedged distribution are much smaller than in the unhedged case which indicates that some part of the uncertainty has been mitigated.

The bar chart in the bottom graphic (Figure 8.5) shows the total uncertainty in the book (as measured by the variance of the simulated 30 year survival probabilities at age 60, allowing for various types of risk sources), before and after applying an index-hedge. The colours represent different sources of risk where the process risk (the pink colour) is the dominant one for the book. By comparing the heights of these two bars, we see that total reduction in variance by hedging the book with an index-based hedge is 80%, which is mainly due to the substantial reduction in process risk.

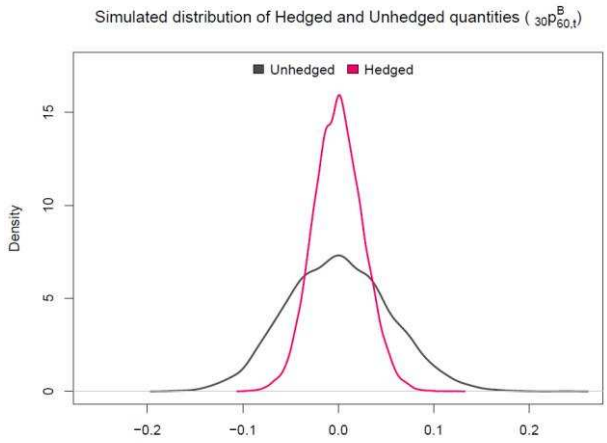


Figure 8.4: Simulated distributions of 30 year survival probability, as measured 10 years into the projection period. 'Unhedged' is distribution relative to the mean survival probability for the book. 'Hedged' is the distribution (relative to the mean) of the difference in survival probabilities between the book and the reference population and is a proxy to the residual risk post application of an index-based hedge

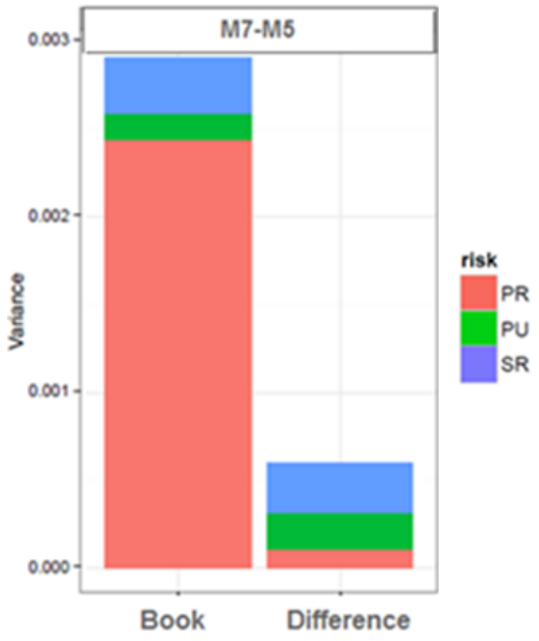


Figure 8.5: Variance of survival probabilities from age 60-90 allowing for process risk (PR), parameter uncertainty (PU) and sampling risk (SR). Recall that PR and PU combined is **demographic risk**.

[48] The variance reduction is defined as: $1 - \frac{Var(_{30}p^{B}_{60,10} - _{30}p^{R}_{60,10})}{Var(_{30}p^{B}_{60,10})}$. It is a crude proxy to the impact of a hedge on the valuation of annuities since annuity rates are closely linked to survival probabilities. However, please note that the derivation and assessment of appropriate basis risk metrics is out of scope of this phase of the research project. Phase 2 of this research (as originally specified) is intended to look at appropriate basis risk metrics.

8.5 Comparison of M7-M5 vs CAE+Cohort Model – Hedge effectiveness results

If there were no need to allow for richer correlations structure in our model, for example if a broad assessment of the residual basis risk associated with a particular contract is required, then the model decision tree would have led us to the CAE+cohort model.

We end this case study by comparing the outputs of these two models when fitted to our sample scheme.

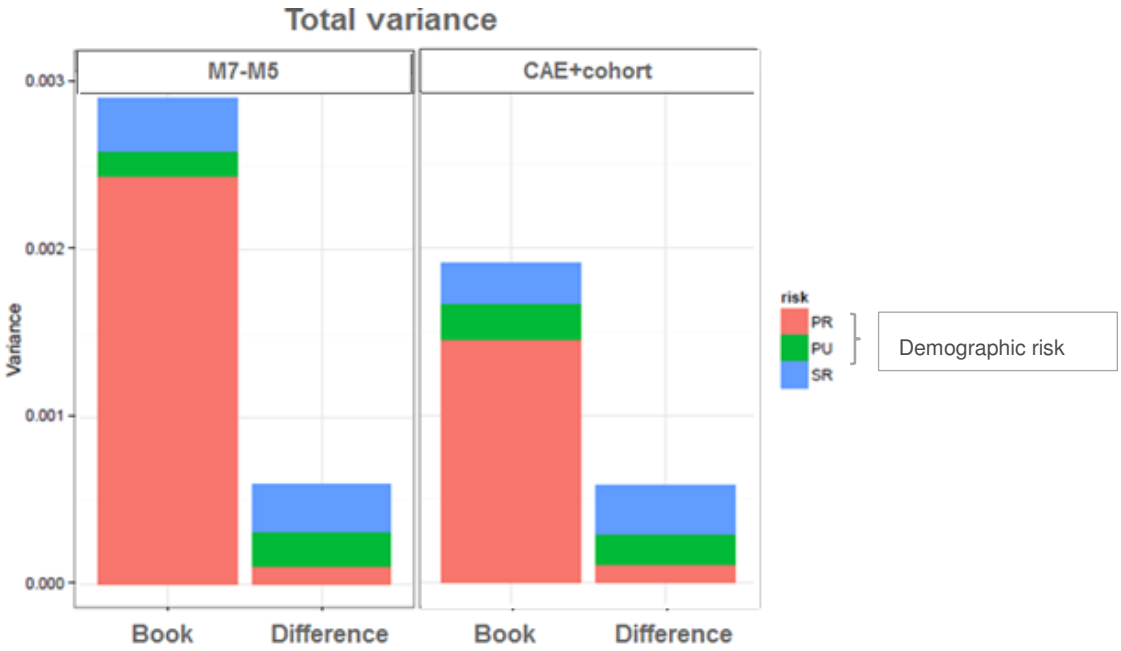


Figure 8.6: Variance of survival probabilities for book and book minus reference under M7-M5 and CAE+cohort models. PR=Process risk; PU=Parameter uncertainty; SR=sampling risk

Focusing on the same variance reduction statistics as used in 8.4, we see in figure 8.6 how:

- both models demonstrate a large reduction in longevity risk when applying an index-based hedge
- the CAE+cohort model projects less uncertainty in the unhedged book than the M7-M5 model due to its simpler structure[49]. This results in a 68% variance reduction from the hedge, compared to around 80% for the M7-M5 model.
- whilst they differ in relative measures (between unhedged and hedged position), they show similar residual risk in the hedged case.

8.6 Summing up

This case study has demonstrated how the flow chart can be used to select a model to assess basis risk and how that model (in this case the M7-M5 model) can be applied. Simulations can be generated, allowing for process, parameter and sampling risk, to assess the benefit of an index-based hedge. Risk reduction metrics can be derived and an understanding of the element of risk being removed (primarily process risk) can be gained. We believe that this methodology is of value to practitioners wishing to form a view on the merits (or otherwise) of index-based solutions.

[49] CAE+Cohort allows for a much simpler correlation structure between annual changes in mortality at different ages than M7-M5 by including a simple time series process. M7-M5 allows for non-perfect correlation between annual changes in mortality at different ages due to the presence of multiple time series.

9 Key challenges and practicalities of direct modelling

9.1 Key challenges of direct modelling

Sections 6, 7 and 8 identified and illustrated the M7-M5 model and (in some situations) the CAE+cohorts model as suggested approaches where the user has sufficient data to support direct modelling. Whilst useful in many situations, it is important for users to consider the limitations, challenges and underlying assumptions inherent in these models when considering their use and drawing conclusions from their output.

A number of key issues to be aware of are discussed below. These challenges are not unique to the M7-M5 and CAE+cohorts models; they are general features which apply to most, if not all, direct modelling approaches.

9.1.1 The past as a guide to the future

When using time series models fitted to historical data there is an implicit assumption that the past is a good guide to the future, both in terms of the:

- 'direction' and 'pace' of travel of mortality improvements within the reference and the book populations; and
- level of future volatility in mortality.

However, akin to the world of financial markets, there are a number of reasons why the past might not be a good guide to the future. In the context of demographic risk reasons could include:

1 The drivers of future changes in longevity may differ from those seen historically as per our discussion in section 3

2 The period used to calibrate the model may – in hindsight - be relatively 'benign', either in the sense of:

  2.1 the drivers of historical longevity trends (e.g. smoking cessation) having impacted society more evenly than future drivers (e.g. affordability of access to anti-ageing therapies) leading us to understate demographic risk.

  2.2 the level of volatility seen between book mortality and reference population mortality, which serves to calibrate the level of future volatility

3 Alternatively, the period used to calibrate the model could – with hindsight – be relatively volatile in terms of book vs reference mortality and so overstate demographic risk

Thus, whilst models fitted to historical data provide an objective starting point for assessing basis risk, we would expect users to wish to interact with the models to incorporate a degree of personal beliefs and / or judgement regarding their wider contextual understanding of the drivers of longevity trends both within their specific book population and the wider population. This is explored further in section 9.2.3.

9.1.2 Default choice of time series

The usual assumption is that the spread between the mortality in the reference and the book will conform to the non-divergence hypothesis in the long run i.e. that the ratio of q^B_{xt}/q^R_{xt} will tend to a constant value as $t \to \infty$. (See for example Li and Lee (2005), Jarner and Kryger (2009), Cairns et al (2011), Li et al (2014) and Zhou et al (2014)).

The non-divergence constraint is commonly captured via the use of a (vector) autoregressive process for the time series indices (κ^B_t) in the book part of the model. This implies that, in the long-run, the spread between the logit of mortality for the book and the reference population will revert from the current level to the historical mean. If the book mortality has moved away from the reference population then it will be projected to converge to the average level of difference over the period to which the model has been fitted. Similarly if historically the

gap between the book and the reference population has closed then it will be projected to widen to a 'steady state' of the average historical difference.

However, using such time series has important implications for the projected mortality trend within the book population and also the difference in mortality between the book and reference population. Specifically:

- it constrains the 'direction of travel' for future mortality differentials
- the variance of the difference in (logit) mortality between the book and the reference population is bounded, limiting the width of the prediction intervals ('funnels of doubt') for the difference in mortality between the book and the reference populations.

In combination these might be considered to understate demographic basis risk and hence overstate the hedge effectiveness. This is of particular significance for users considering longer term hedging instruments (e.g. 20+ years) than the shorter term (typically 10 year) structures which are currently receiving most focus.

No thorough investigation on the implications of this assumption has been performed so far. Detailed guidance on the choice of alternative time series models, allowing or not for divergence, and allowing for the nuances of user judgement, is beyond the scope of this phase of the research. However, the work we have done suggests that this is a very important assumption. This is illustrated in section 11 where we highlight the implications of different choices of time series model within the context of the multi-population modelling required under the characterisation approach.

We, therefore, would encourage users to interact with the choice of time series in order to:

- understand the implicit assumptions; and
- potentially modify the choice of time series process to reflect their own beliefs.

Appendix C sets out a range of time series which could be considered.

9.1.3 Potential overestimation of basis risk for smaller populations

For small book populations, the instability in crude death rates may transfer to the estimates of the time index values. The time series will be more volatile as they implicitly incorporate part of the sampling noise of death rates. Consequently, both parameter and process risk may be overestimated resulting in an underestimation of the hedge effectiveness.

We can see this by contrasting the results from our case study in section 8 with those seen for a smaller scheme with just 5,000 lives and 7 years of back history. Figure 9.1 shows the variance of the difference between the book and reference population in terms of the period survival probability from age 60 to 90, ten years into the future under the CAE+Cohorts approach. This is a broad proxy to the variability of outcomes post an index-based hedge. We can see how each of the elements of risk increases for the smaller book, including the two key components of **demographic risk**, i.e. parameter uncertainty and process risk.

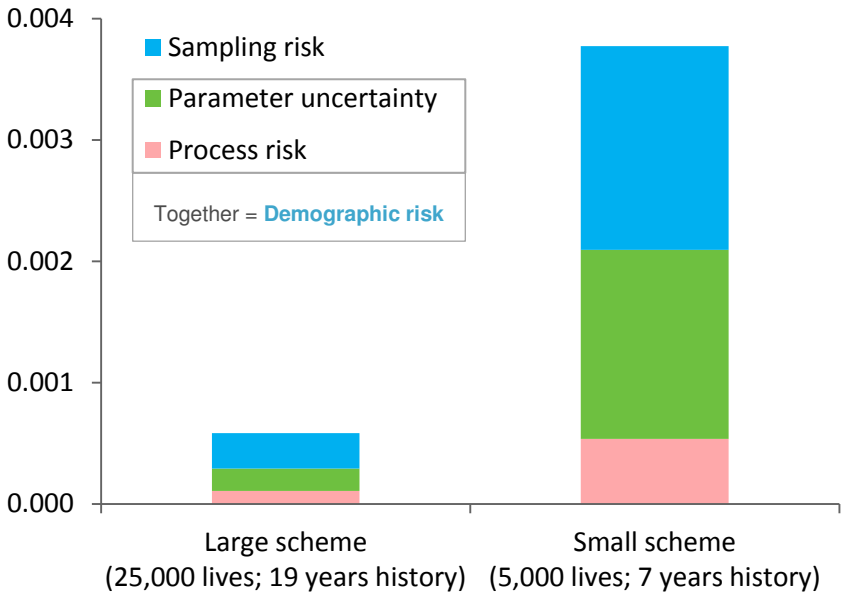


Figure 9.1 Comparison of variance of difference between book and reference population under CAE+cohort for two different book populations.

- **Sampling risk:** Having a smaller book size should lead to a greater variability in actual outcomes arising from sampling risk as seen in the chart.
- **Parameter uncertainty:** By having a smaller number of lives we are less certain about the parameters in our model for the difference in mortality between the book and the population. Hence the value shown is artificially high reflecting the difficulty in evaluating the underlying mortality trends rather than the underlying uncertainty in those trends. For example if we understood more about the underlying nature of demographic risk we may have more confidence in the underlying trends within the book.
- **Process risk:** The small size of the book leads to increased variability in the observed values of the time indices for the difference in mortality between the book and the reference populations. This feeds through to the volatility parameters of the random innovation terms within our time series (the ξ_t^B), and so is also liable to lead to overestimation of basis risk.

It is reassuring that Cairns et al (2014) hint at this issue beginning to be material for book sizes less than about 25,000 lives i.e. a similar threshold to that contained in our decision tree.

For books below 25,000 lives, our approach of focussing on a characterisation approach (rather than direct modelling) for smaller books mitigates this risk. An alternative way to avoid this issue would be to consider a Bayesian approach, as proposed in Cairns et al (2011).

9.1.4 Forecasting horizon

Our research has focussed on assessing the performance for up to a 15 year forecasting horizon. The effectiveness of the models is untested over longer time periods.

Evaluating hedging instruments' effectiveness over longer horizons requires further analysis and care from the user to ensure that the considered models remain appropriate, or to adopt alternatives as necessary.

This issue is of particular importance when considering longer dated instruments or structuring a hedge to provide run off protection.

9.1.5 Older ages

The modelling approaches adopted only fit to and project the range of ages included in the data. Older ages (typically 90+ or 95+) will require some form of subjective (structural) assumption to be applied. Care is needed in doing this, as the approach taken can impact the modelled variability of outcomes for the reference and book populations and hence also the view of demographic risk and hedge effectiveness.

9.1.6 Alternative reference populations

Our model testing is based on using England and Wales data as the reference population. Therefore the results may not directly carry over to other reference populations (especially if materially different in nature, such as from a different country).

In cases where the user is attempting to carry the results of this work across to other reference populations, close attention should be given to the goodness-of-fit of the model for the reference population. If the results are unsatisfactory then further analysis and model testing may be required. Nevertheless, the research approach we have taken should be replicable for other countries and we would expect many of the key conclusions still to hold.

9.1.7 Hedging non-UK portfolios

In our specification and analysis of the two population model for the reference and the book populations, we have implicitly assumed that the book is a subset or is closely related to the reference population on which the index is based.

Other cases (hedging UK mortality with another country, say US, mortality) would require a deep understanding of the differences between the two countries' mortality. Such differences may not be captured by the structure of the two population models we have proposed and the relative approach we have pursued may have to be substituted by a simultaneous modelling of the two countries' mortality.

9.2 Addressing the practicalities of direct modelling

Sections 6, 7 and 8 identified and illustrated pure direct modelling approaches with a focus on men. In practice users may wish to combine stochastic direct modelling techniques with their own views on certain aspects of longevity (particularly for the book) and will be concerned with portfolios containing both men and women.

9.2.1 Allowing for known base rates

Both of the direct modelling approaches suggested in our decision tree produce base rates for the book population. In practice, however, the user may have an alternative approach to the base rates which he or she would prefer to use – for example due to an established mechanism for assessing this, or due to a more granular method such as individual underwriting. This may lead the user to having base rates $q_x^{B,user\ base}$ which are to apply at time T.

9.2.1.1 Direct substitution

In some circumstances it is possible to substitute the user's base rates into the fitted model prior to projection of future mortality rates.

For example under the CAE+cohorts approach the α_x^B has a natural interpretation as the difference in mortality levels between the book and the reference population. This means that it is easy to superpose the user's preferred base rates by:

1 Substituting the fitted α_x^B with calculated $\alpha_x^{B,user}$ for each age x :

$$\alpha_x^{B,user} = logit\left(q_x^{B,user\ base}\right) - \alpha_x^R - \beta_x^R(\kappa_T^R + \kappa_T^B) - \gamma_{T-x}^R$$

2 Modifying the identifiability constraints used to fit the model to enable the direct substitution. Specifically: remove the constraint defined by equation (7.13) as by fixing α_x^B changes in levels of κ_t^B are no longer permitted.

9.2.1.2 Use of reduction factors

Some models, such as the M7-M5 model, do not support direct substitution without modification to the model (e.g. as per Plat (2009a)) and overcoming the associated identifiability challenges (see Hunt and Blake 2014b). Here, an alternative approach is to:

1 Calculate the implied annual improvements (reduction factors) from the simulation of future mortality for the book population:

$$RF_{x,t}^{B,fitted} = \frac{q_{x,t}^{B,fitted}}{q_{x,T}^{B,fitted}}$$

2 Apply these to the users own base table to obtain revised simulations for the book population:

$$q_{x,t}^{B,user} = q_{x,T}^{B,user} \; RF_{x,t}^{B,fitted}$$

9.2.2 Modelling men and women

In practice most book populations contain a mix of men and women and so the user may want to model men and women as part of assessing hedge effectiveness. A number of possible approaches exist, with the choice depending on the type of hedge that is being considered and whether the available indices are gender based (the usual case) or unisex.

9.2.2.1 Hedging with separate indices for men and women – modelling men and women separately

If hedging is to be done using two sets of indices, one for men and one for women, then direct modelling of reference and book populations (as illustrated in section 8) could be undertaken separately for both genders, *provided* the data for each gender met the number of lives / back history criteria.

The primary challenge would then be to allow for the correlation between the genders (at both reference and book level). This can be overcome by either:

- modelling mortality for men and women simultaneously (at least for the reference population), for instance with the use of a multivariate time series model for the period indices of men and women (see Li and Hardy (2011)); or
- modelling mortality for men and women independently and aggregating the results with an adjustment to allow for correlation (similar to what is done when aggregating risks in, for example, the framework of Solvency II)

9.2.2.2 Hedging with separate indices for men and women – unisex modelling

A challenge of modelling men and women separately is that book data volumes may be too low to allow separate direct modelling for men and (particularly) women. Most books will be below 25,000 lives in aggregate, and (almost) certainly will be below this for one or other gender.

One possible approach is to treat the book as a single entity (notwithstanding the fact that it is a mix of men and women) and model the reference population using an appropriate blend of male and female reference populations. This would allow the direct modelling approach illustrated in section 8 to be applied.

Care would be needed to weight the historical reference population data in line with appropriate male / female weights e.g. in line with the mix of male / female hedging instruments to be used.

9.2.2.3 Hedging with a single index – modelling men and women separately

In some circumstances, only a single index may be available – or indeed the user may elect to use a single index on grounds of broader considerations such as liquidity. For example, with the majority of UK pension liabilities linked to the longevity of men it is conceivable that we will see greater liquidity developing for a male index.

One approach would be to model the men and women within the book population separately. Taking the example where hedging involves only a male index, the output of interest would be simulations for both male and female books but only the male reference population. Nevertheless, it is highly likely that fitting and simulation of the female reference population would be required in order to model the female book population mortality relative to that reference population.

This approach is akin to that described in section 9.2.2.1 with one important caveat. Modelling men and women independently and aggregating the results with an adjustment to allow for correlation is unlikely to be suitable, as it would not capture the basis risk between male reference population and female book. It would therefore be necessary to model mortality for men and women simultaneously (at least for the reference population), for instance with the use of multivariate time series models.

9.2.2.4 Hedging with a single index – unisex modelling

Alternatively, modelling the book population on a unisex basis avoids the need for simultaneous modelling of men and women. Here, the unisex approach to the book population described in section 9.2.2.2 could be applied. The reference population would be male only, and the book population would be the combined (male and female) book.

Care may need to be taken in relation to any time trend in mix of men and women in the book (akin to the issue of a changing socio-economic mix over time) and the weight of liabilities for men and women in order to ensure that the resulting assessment of hedge effectiveness is appropriate.

9.2.3 Applying user judgement

The most likely area where the user will wish to apply judgement is the extent to which he or she relies on the past as a guide for the future, and thus amends the parameters of the processes used for the projection of the time series and/or the nature of the times series used to reflect this view.

Users may also wish to move away from the usual VAR(1) assumption for the time indices given the embedded assumption that in the long-run the spread in mortality between the book and the reference population will converge to its historical mean and the limited width to the funnel of doubt for this difference.

Appendix C describes a selection of possible time series. The appropriate choice will very much depend on a combination of what the book data supports and the user's beliefs. As a very high level guide:

- **Capturing linear trends in historical data:** Use time series such as multivariate random walk with drift (MRWD) when the user prefers unbounded prediction intervals ('funnels of doubt') or vector autoregressive processes (VAR) about a linear trend for bounded variability and prediction intervals.
- **Trending to a view of stable relative mortality:** Convergence to a pre-specified long-term level input by the user can be achieved by – for example – constraining the parameters of the AR(1) process in line with Börger et al (2013)
- **More complex evolutions of book vs reference population:** In principle more complex structures can be embedded in the time series structure. Taking the VAR processes described in Appendix C as an example, this would mean replacing the linear trend term ($\Phi_{01}t$) with an alternative more complex

function of time. Alternative co-integration or vector error correction models could be used (see e.g. Yang and Wang (2013) and Zhou et. al (2014)).

Users may also wish to combine some of these options; for example starting with some linear trends capturing narrowing or widening of differentials in the shorter term before trending to some stable relative rates. A detailed description of how users might achieve this is beyond the scope of this stage of our research – however, we would encourage further research into appropriate choices of times series and the possible ways of embedding judgement into the modelling.

10 Characterisation Approach

In practice, most pension scheme books and life company portfolios will fall below the size requirements for direct modelling – having fewer than 25,000 annuitants (even across men and women) or less than 8 years of reliable back history. Most users will, therefore, find themselves requiring an alternative approach if they are to avoid overestimating demographic basis risk as per 9.1.3.

Even where the book is sufficiently large with long enough experience history to use direct modelling, an alternative indirect approach is still likely to be useful; either as a means of a pragmatic initial assessment of the quantum of basis risk, or as an alternative approach as part of considering model risk.

This section explains the construction and implementation of a characterisation approach. Case studies are provided in section 11.

10.1 How the characterisation approach works

Instead of using the experience data of the book itself, the basic principle of the characterisation approach is to map the book onto a small number of characterising groups which:

- capture the majority of the source of demographic risk
- can be projected using an alternative data source with a more reliable and longer back-history of mortality experience

Schematically, this approach can be thought of as:

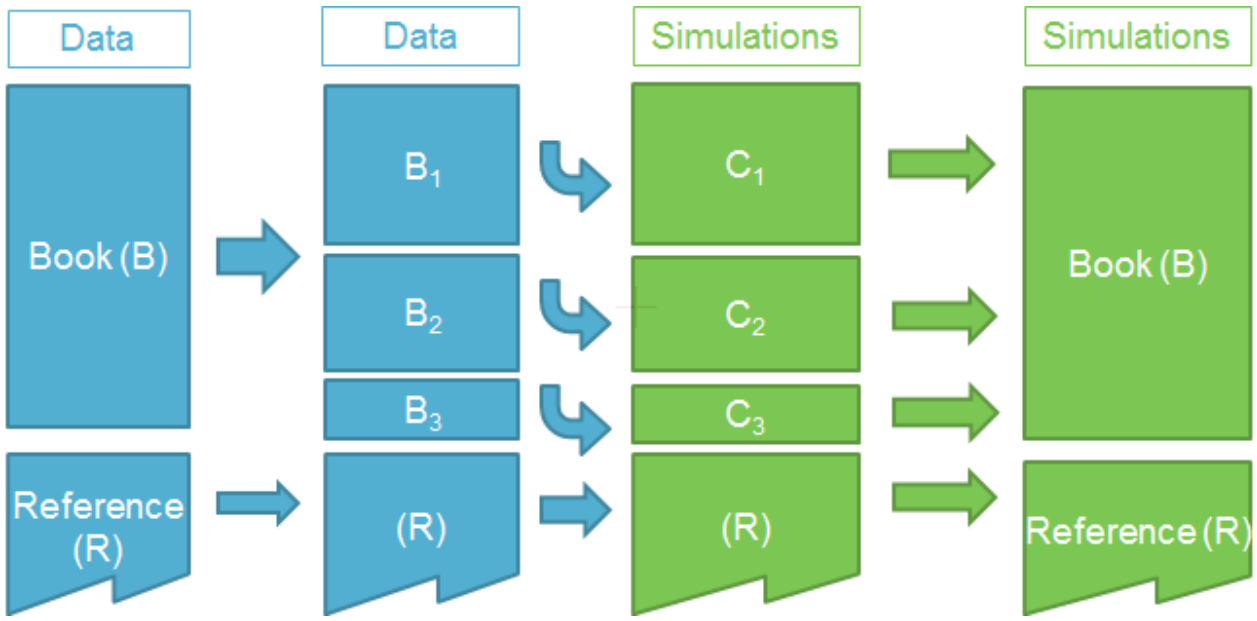


Figure 10.1: High level schematic of characterisation approach

The modelling process takes two stages:

Stage 1: Pre generate simulations for a set of characterising groups (C_1,C_2,…) and the reference population (R)

This need only be done once and can be done quite separately to the application to the specific book. These simulations will include both parameter uncertainty and process risk.

Stage 2: Create simulations for the book and reference population
This involves:

- Segmenting the book into groups $B_1,B_2,\ldots$ using the same criteria as define the characterising groups
- Obtaining the pre-generated simulations of mortality rates applicable to each characterising group and the reference population for each age and each future year
- Simulating survivorship of each segment of the book $(B_1,B_2,\ldots)$ using binomial sampling and the pre-generated simulated mortality rates for the corresponding characterising group $(C_1,C_2,\ldots)$
- Adding together the simulations for each segment to get a simulation for the total book population
- Using that simulation of the book population with the corresponding (pre-generated) reference population simulation to compare index and book outcomes

These steps result in the large number of simulations required to form the assessment of basis risk.

10.2 Key questions within characterisation approach

In order to apply the method outlined in section 10.1 five key questions need to be answered:

1 What dataset ('characterising population') should be used for the pre-generated simulations?

2 How should the characterising population be segmented into 'characterising groups'?

3 What stochastic model (and time series) should be used to simulate these groups?

4 How should the book population be simulated given these pre-generated simulations?

5 What adjustment, if any, should be made for potential residual basis risk? (i.e. the extent to which the characterising groups may not capture all of the underlying basis risk)

Sections 10.3-10.7 look at each of these decisions in turn using example characterising populations to illustrate the thought process that is involved in each decision.

10.3 What dataset to use for characterising population?

The dataset used for the characterising population needs to be large and have sufficiently long back history in order for it to give reliable simulations. Specifically, it will ideally be an order of magnitude greater than the minimum threshold for direct modelling, i.e. of the order of 250,000 lives, so that it can support direct modelling on the characterising sub-populations.

In order to be useful for the characterisation approach, the dataset needs to have sufficient information to allow segmentation into sub-groups that are likely to capture future longevity variations. In addition the variables used for segmentation must be available and have a consistent definition / meaning within the book populations. Example variables could be postcode or pension/annuity income. However in the latter case care would be needed to understand whether the characterising population related to defined benefit pensioners / bulk-purchase annuitants or individual annuitants as pension potentially has a different meaning between these groups.

In a similar vein it is important that the characterising population is relevant to the book population in order to reduce the scope for material residual basis risk not captured by this approach. This means that it would be quite reasonable – and indeed where the book population is a select group of lives desirable – for the characterising population to differ from the reference population. In the case of using an England & Wales reference population, possible options for the characterising population include:

- ONS data (split for example by a socio-economic variable such as postcode based index of multiple deprivation);

- CMI dataset (e.g. the SAPS data which has a back history of experience data split by pension amount); and
- Club Vita dataset (which can be split by a range of affluence and postcode metrics).

In the context of this report, we have chosen to illustrate only one of the ONS and CMI dataset since each can only be used to create characterising groups using one variable. The more limited back-history (currently 2000-2012[50]) combined with the year to year variation in contributing schemes for the CMI SAPS dataset pose challenges in calibrating the time series models, and hence we focus on the ONS dataset. The ONS data contains exposure and death data for individual calendar years and five year age bands split by Index of Multiple Deprivation (2007 base). This dataset is restricted to England by necessity (as the IMD weights several components each of which are expressed relative to the national average).

We also illustrate the considerations in choosing and projecting the sub-populations for the Club Vita data. This is informative in two regards. Firstly, it enables a discussion of determining the characterising populations when we have multiple dimensions available for use in the segmentation. Secondly, the historical data support the use of different time series to the ONS data enabling us to illustrate some important considerations in respect of the choice of time series.

10.4 Choosing the characterising groups

Having chosen our characterising population we need to identify how to segment it into groups which we believe will capture most of the heterogeneity in future longevity trends and thus demographic risk. A natural starting point in this regard would be differences in historical improvements, although the user may also wish to keep certain groups separate where he or she has a particular belief regarding the potential for divergent trends between those groups.

We suggest six core principles that should be applied and balanced when choosing the characterising groups:

1 **Credible size:** Each resulting group needs to be sufficiently large that we can confidently apply direct modelling to it, but not so large that it dominates the assessment of basis risk for individual books.

2 **Separate clear differences in improvements:** We wish to ensure that the groups capture the major differences seen in historical improvements between different parts of the characterising population.

3 **Group where similar improvements:** Where particular parts of the characterising population have experienced similar levels of improvement we would generally keep these together.

4 **Separate clear differences in mortality levels:** Where different groups of the characterising population have very different current levels of mortality we would wish to keep these separate as they are liable to be subject to different major causes of death and so respond differently to future longevity improvements (even if they have exhibited similar trends in the past).

5 **Interpretable:** The resulting groups should contain like individuals (i.e. similar in terms of real world features such as affluence) and thus have some interpretable and intuitive meaning. This enables the user to apply their broader understanding of the drivers of improvements in exercising judgement within the modelling of these groups.

6 **Manageable number:** The resulting number of groups should not be very large as this will materially increase the number of parameters in the multi-population model which will be used to simultaneously simulate the reference population and each of the characterising populations. This in turn will magnify the parameter uncertainty. Equally, the number of groups should not be too small as otherwise we will inadequately characterise the demographic risk. For the purposes of illustration we have not sought to

[50] Working paper 73 covers data over the period 2005 to 2012; earlier Working Papers cover data back to 2000.

determine an optimal number; and have instead illustrated the approach to achieving the other principles based upon having 3 clustering groups.

We illustrate the application of these principles to the ONS and Club Vita characterising populations below.

The details of the application of the principles to the Club Vita dataset are covered in Appendix D.

10.4.1 One dimension (ONS IMD data for English men)

The ONS IMD data enable us to group a characterising population of the English population by deprivation quintile. Here we are looking to reduce the five groups available in the data into three characterising groups. Whilst this can be done using the kind of statistical techniques described in 10.4.2, in these circumstances a pragmatic application of the principles may be preferable.

By virtue of each quintile capturing 20% of the national data we can be confident that any grouping we produce will meet the credible size requirement; and, thus, our focus is on principles 2-4 above.

From Figure 2.2 in section 2.1 (repeated to the right) we can immediately see how:

- There are clear differences in improvements for the most deprived quintiles (Q4 and Q5) compared to the other quintiles
- Q1, Q2 and Q3 have similar levels of historical improvement and so potentially can be grouped together

Looking at the levels of mortality for each group we find that there are clear differences in mortality levels with Q3 much higher than Q1 and Q2 (see Villegas and Haberman (2014)). The potential for the drivers to be different for Q3 therefore suggests that this quintile is treated as a separate group.

Bringing this together gives the following three groups, which clearly also adhere to the interpretability principle:

- Below average deprivation (Q1-Q2)
- Average deprivation (Q3)
- Above average deprivation (Q4-Q5)

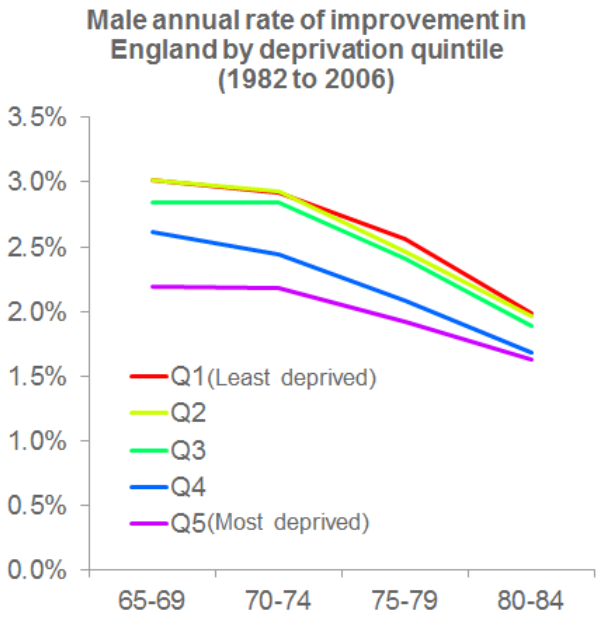


Figure 10.2: Annualised improvements in mortality, England by deprivation quintile. Based on Table 1 and 2 in Lu et al. (2013)

10.4.2 Multiple dimensions available (Club Vita data for men)

The Club Vita DB annuitant dataset enables us to group individuals by a wide range of characteristics which could be indicative of differences in future longevity trends and thus demographic risk. In Section 2.1 we saw that – of those characteristics which are widely available within pension schemes / BPA books – combining pension income and postcode based deprivation measures provided the most parsimonious fit to historical experience.

We create our characterising groups via the following steps:

1 Segment the spectrum of values each variable can take into discrete groups, and so in turn the characterising population into 'cells'

2 Identify a distance metric which measures the level of dissimilarity between these cells striking a balance between our competing principles

3 Use statistical techniques (specifically partitioning about medoids and fuzzy analysis) to cluster these cells into our desired number of groups

4 Interpret the results of the clustering and consider whether appropriate to adjust to ensure groups are both interpretable and credible in size

The detail of applying these steps is covered in Appendix E and leads to three illustrative groups which could be described as the **'modest means'**[51], **'middling'** and the **'higher wealth'**.

Pen/IMD	Q5 (most)	Q4	Q3 (mid)	Q2	Q1 (least)
<5k					
5-10k					
10k+					

It is reassuring that the resulting characterising groups meet our principle of having credible data volumes – having comfortably in excess of the 25,000 lives required for direct modelling:

	Modest means	Middling	Higher wealth
Number of lives (2010)[52]	61,879	81,883	90,722

We can also see from Figure 10.3 that the resulting groups have captured meaningful differences between the groups both in terms of the mortality levels (which have the expected ordering) and mortality improvements (which for the example age range[53] of 75-84 exhibit a 'smile' shape).

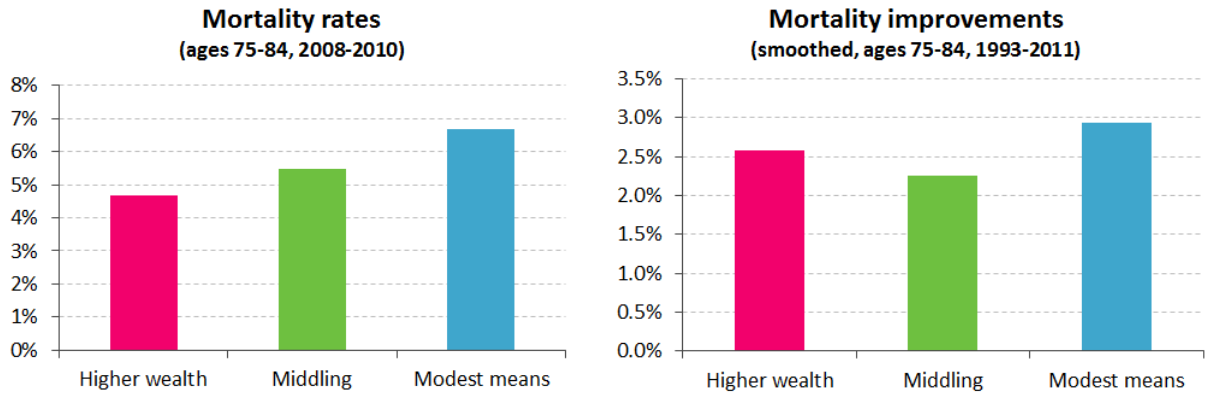


Figure 10.3: Mortality rates and smoothed annual improvements for characterising groups based on Club Vita data (ages 75-84)

[51] Modest means in the sense that low pension and high deprivation geography liable to lead to modest non-pensions wealth

[52] Men aged 65-90 with good quality pension and postcode data (and post application of scheme level exclusions for biases between clean death and exposure data as per Appendix A)

[53] A curtailed age range is used to reduce any confounding arising from changes in the average age of each group over time.

Looking across pension schemes within the Club Vita dataset we can also note how the groups are likely to be informative in capturing differences in exposure to longevity trends and demographic risk. Whilst the typical scheme[54] has a broadly even split of membership across the groups (on a lives basis), some schemes are noticeably skewed either towards the 'healthy wealthy' groups or the 'unhealthy poor' groups.

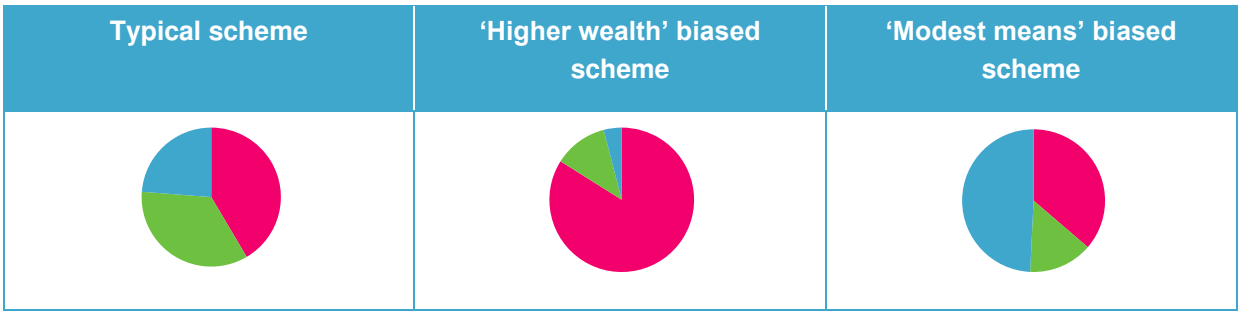


Figure 10.4: Distribution of membership by our illustrative characterising group for different schemes within the Club Vita dataset.

Within any specific book population the financial exposure to demographic risk will be skewed towards the 'healthy wealthy' group compared to the lives distributions shown above. In light of this feature, some users may wish to further split the wealthy group, particularly if, on an amounts basis, the liabilities are almost exclusively within that group.

10.5 Which model to use

Having identified the characterising groups we need to simulate these populations, simultaneously with the reference population. By construction, each of the characterising groups is large enough so that, if we were to treat it as a book population, then under the flowchart introduced in section 7 we would use the M7-M5 model to simulate it alongside the reference population. A natural approach therefore is to use a multi-population extension of the M7-M5 whereby each of the characterising groups is modelled using the M5 formulation but where an allowance is made for correlation between the characterising groups.

10.5.1 Multi-population M7-M5

10.5.1.1 Reference population

Using the notation introduced in section 4 we have:

$$D_{xt}^{R} \sim Bin\left(E_{xt}^{R}, q_{xt}^{R}\right)$$

$$\text{logit}\, q_{xt}^{R} = \kappa_t^{(1,R)} + (x-\bar{x})\kappa_t^{(2,R)} + ((x-\bar{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma_{t-x}^{R}$$

The vector of time series indices $\boldsymbol{\kappa_t^{R}} = \left(\kappa_t^{(1,R)}, \kappa_t^{(2,R)}, \kappa_t^{(3,R)}\right)^T$ are then modelled as a MRWD as per section 5.2 i.e.

$$\boldsymbol{\kappa_t^{R}} = \boldsymbol{d} + \boldsymbol{\kappa_{t-1}^{R}} + \boldsymbol{\xi_t^{R}}, \qquad \boldsymbol{\xi_t^{R}} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma^{R}}),$$

If it is necessary to extend the cohort term γ_{t-x}^{R} outside of the birth generations included in the dataset then we follow an ARIMA(1,1,0) approach as per section 5.2:

$$\Delta\gamma_c^{R} = \phi_0 + \phi_1 \Delta\gamma_{c-1}^{R} + \varepsilon_c^{R}, \qquad \varepsilon_c^{R} \sim N(0, \sigma_R^2)$$

where $\Delta\gamma_c^{R} = \gamma_c^{R} - \gamma_{c-1}^{R}$.

[54] As measured by the mix in the aggregate dataset

10.5.1.2 Characterising groups

Denoting by a superscript of C_i the quantities for each of the characterising groups we have:

$$D_{xt}^{C_i} \sim Bin\left(E_{xt}^{C_i}, q_{xt}^{C_i}\right)$$

$$\text{logit}(q_{xt}^{C_i}) - \text{logit}(q_{xt}^{R}) = \kappa_t^{(1,C_i)} + (x - \bar{x})\kappa_t^{(2,C_i)}$$

We then need to model the multivariate time series κ_t^C:

$$\boldsymbol{\kappa_t^C} = \left(\kappa_t^{(1,C_1)}, \kappa_t^{(1,C_2)}, \ldots, \kappa_t^{(1,C_n)}, \kappa_t^{(2,C_1)}, \kappa_t^{(2,C_2)}, \ldots, \kappa_t^{(2,C_n)}\right)^T$$

in an appropriate way. Possible multivariate time series which will embed a correlation structure include multivariate random walk with drift and vector-autoregressive processes (see Appendix C).

10.5.2 Choice of time series

Similar to the direct modelling situation, the choice of appropriate time series depends on both the trends in the historical data for the characterising groups and the user's beliefs / judgements. As such it is not appropriate to give a general time series formula, although we discuss the key issues that are likely to need to be addressed below by looking at the $\kappa_t^{(1,C_i)}$ time indices for our two illustrative characterising populations. We return to the other time series indices in our case study (section 11).

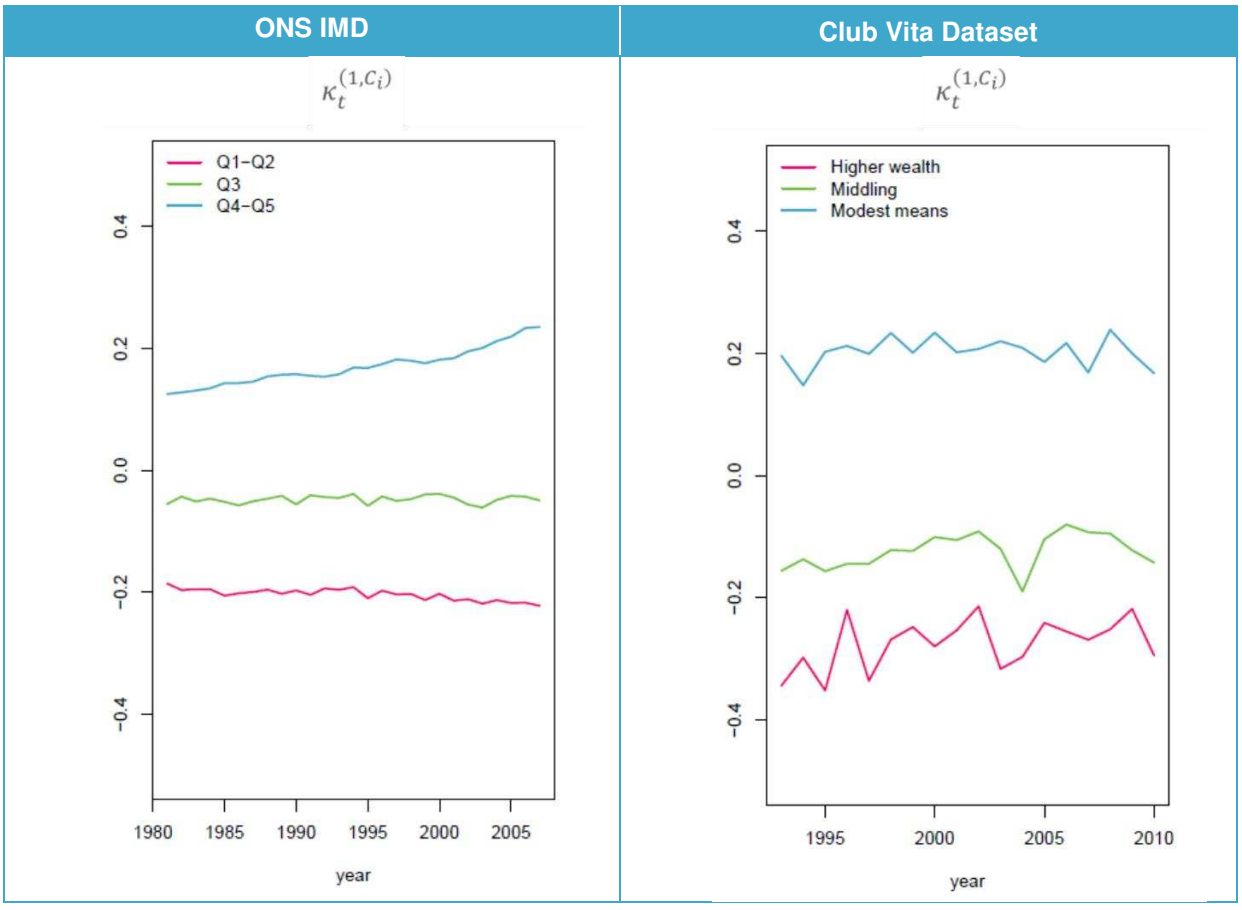


Figure 10.6: Fitted $\kappa_t^{(1,C_i)}$ terms for two different characterising populations (ONS IMD and Club Vita)

10.5.2.1 ONS IMD

We can see how:

- $\kappa_t^{(1,Q4-Q5)}$ is positive with a clear upward trend, i.e. average levels of mortality are rising relative to the reference population (i.e. the gap in mortality is widening vs England & Wales)
- $\kappa_t^{(1,Q1-Q2)}$ is negative and shows a slight downward trend, i.e. average levels of mortality are falling relative to reference population (so gap widening vs England & Wales)

A natural inclination would be to model continuation of these trends – this could be achieved via time series such as:

- Multivariate random walk with drift (MRWD)
- VAR with linear trend

In each case this would assume a perpetual widening of the gap in relative mortality as measured by the logit of mortality[55]. Alternative assumptions could be:

- A stable gap in mortality differentials at the last known value (VAR with a known constant)
- Convergence to the average level of differential over the time period fitted to (VAR about constant)
- A continued widening of the gap before a closing of gap (e.g. VAR with appropriate function of time for the trend)

10.5.2.2 Club Vita data

We can see how:

- there is less clear evidence of strong trends for the Club Vita data; and
- the 'higher wealth' group shows some evidence of an upward trend.

We could therefore use a multivariate random walk with drift or VAR with linear trend to capture the slight trends in the historical data. If we do this, then care is needed as the trend in the 'higher wealth' group is liable to be sufficiently strong that it will ultimately catch up, and overtake the other groups leading to counter-intuitive relative levels of mortality.

This can be avoided by using alternative time series which:

- limit the period for which drift occurs to a specific time period; or
- use a more complex time function to dampen, cease or reverse the trend over time; or
- have no trend term for example VAR with constant.

10.6 Simulating the book population using the characterising groups

Having generated simulations of the q_{xt} for the characterising groups and the reference population – allowing for parameter uncertainty and process risk – we then need to apply these to the book population including capturing the sampling risk therein.

First, we need to map the book population on to the characterising groups. In general this will be a simple task. Inevitably, though, some individuals in the book may have missing or unreliable information in the rest of the

[55] Note that this need not feed through to a widening of the gap in life expectancy though. In general if we have two groups of individuals, then the group with the highest mortality levels will see slightly larger increases in life expectancy for the same level of year on year reductions in mortality. Thus the widening of gap in logit mortality may not be sufficient to cause a widening of gap in life expectancy.

variables used in the characterisation. For example they may have an overseas postcode (which for the purposes of IMD can include Scotland & Wales[56]). Provided the proportion of individuals whose data does not enable them to be mapped to a characterising group is modest, a pragmatic approach can be adopted. This would typically involve assigning these individuals in accordance with the split between characterising groups seen in the book as a whole (with the results rounded to the nearest integer as we must have an integer number of lives). Where the characterising groups are defined across multiple dimensions it would be desirable to apply this allocation using the splits seen between groups seen in the book conditional on any known values. For example, if we had an individual with a pension of £7,000 p.a. but without IMD information then under the Club Vita characterising groups we know this individual cannot be in the **'Middling'** group and we should reflect this when assigning his exposures.

Having calculated the number of individuals currently alive at each age in each characterising group we then simulate each sub-group separately using binomial sampling as illustrated below[57]:

	Start	Simulated period				
Age/year	2010	2011	2012	….	2039	2040
60	$N^{C_i}_{60,2010,j}$					
61	$N^{C_i}_{61,2010,j}$	$N^{C_i}_{61,2011,j}$				
62	$N^{C_i}_{62,2010,j}$	$N^{C_i}_{62,2011,j}$	$N^{C_i}_{62,2012,j}$			
⋮	⋮	⋱	⋱	⋱		
89	$N^{C_i}_{89,2010,j}$	$N^{C_i}_{89,2011,j}$	$N^{C_i}_{89,2012,j}$	…	$N^{C_i}_{89,2039,j}$	
90	$N^{C_i}_{90,2010,j}$	$N^{C_i}_{90,2011,j}$	$N^{C_i}_{90,2012,j}$	…	$N^{C_i}_{90,2039,j}$	$N^{C_i}_{90,2040,j}$

Figure 10.7: Simulating the survivorship of the initial membership within each characterising group

where $N^{C_i}_{x,t,j}$, the number of individuals in characterising group C_i within the book who are alive aged x at time t under simulation j, is drawn as a binomial simulation from $Bin(N^{C_i}_{x-1,t-1,j}, 1 - q^{C_i}_{x-1,t-1,j})$.

Finally, for each age and each future point in time we can sum the numbers alive within the book for each characterising group to get a simulation for the book population.

10.7 What adjustment (if any) should be made for residual demographic risk?

In applying the characterisation approach we are assuming that we can effectively capture all the potential for demographic risk via a small number of characterising groups. Although the groups have been chosen so as to distinguish individuals in terms of mortality improvements there will be demographic risk which cannot be captured. Thus, the 'true' level of basis risk is likely to be a little higher than implied under this approach. In this regard it is reassuring that our (limited number of) case study book populations in section 11 show modest differences between the hedge effectiveness calculated by the characterisation approach and direct modelling, with the characterisation approach giving variance reduction statistics up to 6% higher[58]. However, as we will

[56] The IMD values as publically available apply are not comparable across nations in the UK. However it is relatively easy to compute a pan UK IMD value – see section 12 for further discussion.

[57] This differs to the approach followed in sections 6 and 8. Specifically, here we assume that the book is modelled under a run-off basis and thus decrease over time. This replicates the application of index-based hedges. In our earlier analysis we were assuming that the size of the scheme was kept constant at the size in the last year of observation to facilitate comparisons between models.

[58] Where comparable time series are used

also see in Section 11, it need not be the case that the characterisation approach suggests higher variance reduction / hedge effectiveness than the direct modelling approach.

11 Case study – characterisation approach

Section 10 gave a detailed description of the characterisation approach which will most likely become practical for users with books/portfolios smaller than required for direct modelling. For a user to apply the procedure to his book/portfolio a number of modelling assumptions need to be made, such as which characterisation population and which time series to use. Section 10 has provided the user with a general guide to these questions along with considering a few alternatives.

This case study of the characterisation approach will demonstrate the actual modelling process, and the results of applying the characterisation simulations when applied to five different pension schemes from Club Vita. Two different characterising populations (ONS data and Club Vita data for men only) will be used to illustrate the approach using the M7-M5 multi-population model.

11.1 Example A: ONS IMD Data

As highlighted in section 10.5.1.1, a natural choice of time series to model the historical trends in $\kappa_t^{(1,C_i)}$ when fitted to the ONS data (split into three socio-economic groups) was either:

- Multivariate random walk with drift (MRWD); or
- VAR with linear trend

By choosing these time series for the two time dependent variables in the M7-M5 model we assume a continuation of the historical trend when projecting as per the funnels of doubt for $\kappa_t^{(1,C_i)}$ and $\kappa_t^{(2,C_i)}$ in figure 11.1.

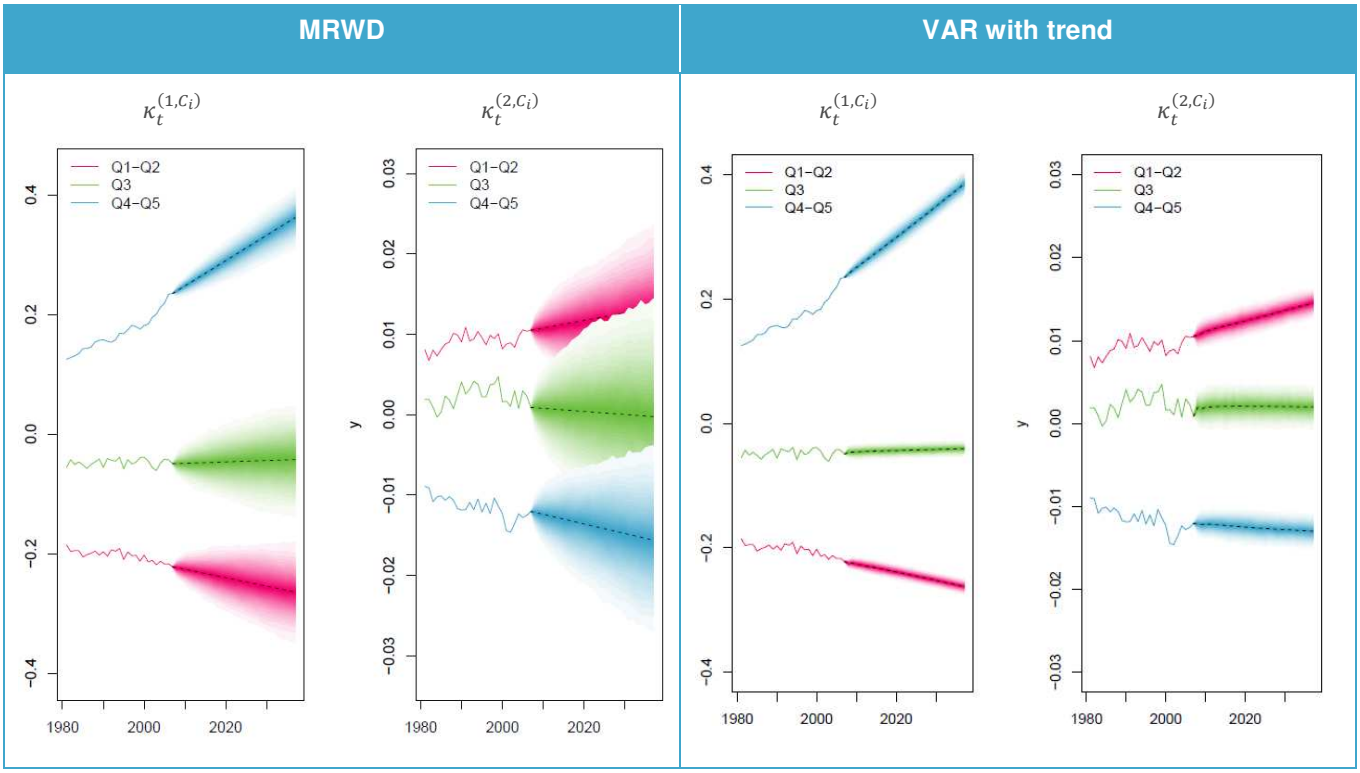


Figure 11.1: Future projections of the fitted parameters for $\kappa_t^{(1,C_i)}$ and $\kappa_t^{(2,C_i)}$, using the ONS data split by three IMD clusters.

Note the reverse order of the clusters in the graphs above, independent of which time series is applied. For $\kappa_t^{(1,C_i)}$ the least deprived group (Q1-Q2) has the lowest value in relative mortality (on the logit scale) and the most deprived group (Q4-Q5) the highest but for $\kappa_t^{(2,C_i)}$ the opposite ordering occurs. This is consistent with the "compensation law of mortality" (Gavrilov & Gavrilova (1991)) whereby mortality converges at older ages; thus increasing/decreasing trends (away from the reference population) in $\kappa_t^{(1,C_i)}$ need to be levelled off with inverse

patterns of $\kappa_t^{(2,C_i)}$ to ensure that the book population converges to the reference (England & Wales) as age increases.

The main reason for considering MRWD and VAR with trend as appropriate time series for the ONS data is to make sure that the clear historical trend would be continuously modelled when forecasting. However, the width of the prediction intervals under the VAR with trend model look unreasonably tight. In contrast, when using the MRWD to model the time series the prediction intervals look more natural. For this reason, our case study models the time depending variables $\kappa_t^{(1,C_i)}$ and $\kappa_t^{(2,C_i)}$ using the MRWD time series.

11.1.1 Simulations under MRWD approach

One way of assessing the reasonableness of the resulting simulations is to look at the simulated life expectancies.

We start by looking at the distribution of simulated period life expectancies at age 65 in 2010 in each characterising group in the ONS data. Since 2010 is the last year of observed data, these distributions reflect parameter uncertainty only, and consequently have very tight densities.

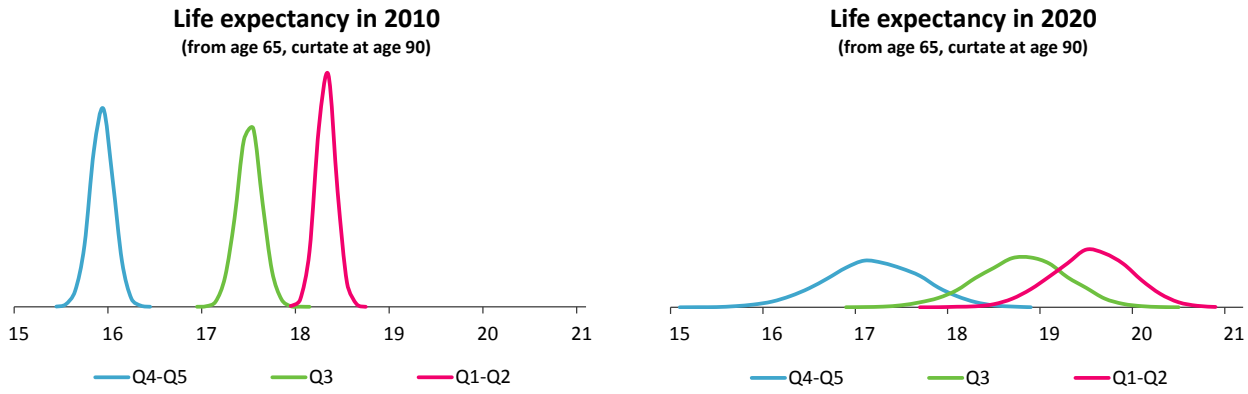


Figure 11.2: Life expectancy projections for ONS IMD characterising groups.

Looking at the simulations in 2020, we observe a much wider spread of outcomes in the life expectancy distributions compared to 2010 since we now have ten years of compounded process risk.

Although the densities overlap come 2020, this need not suggest that the projections are incoherent i.e. the life expectancy of lower socio-economic groups overtakes those of higher socio-economic groups. The correlated nature of the projections for the individual characterising groups means that simulations to the left (right) of the bell curve for one group, will also tend to be to the left (right) of the bell curve for the other groups. Indeed on closer inspection of the projected differences in life expectancy, we find a very small probability that for example Q3 outlive Q1 & Q2 (0.1%) or that Q4 & Q5 are outliving Q3.

We can also see that all groups experience a broadly similar increase in life expectancy between 2010 and 2020, despite the divergence happening in underlying mortality differentials (chart to right). This reflects the mortality / longevity dynamics whereby a smaller percentage reduction in mortality is required amongst the shorter lived groups for the same absolute increase in life expectancy.

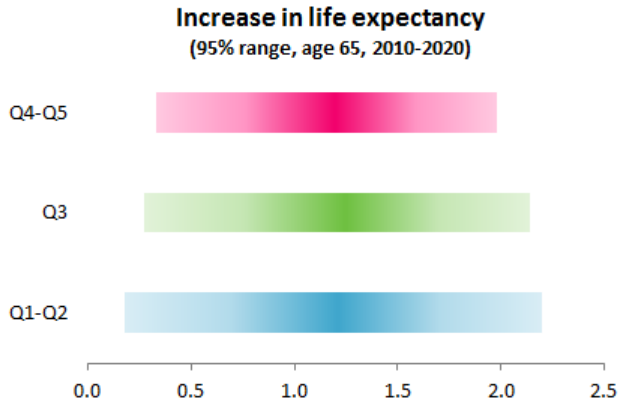


Figure 11.3: Projected increase in life expectancy for ONS IMD characterising groups, 2010 to 2020.

11.2 Example B: Club Vita data

In section 10.5.1 we have observed less clear evidence of strong trends (relative to the ONS data) in the historical values for $\kappa_t^{(1,C_i)}$ and $\kappa_t^{(2,C_i)}$ when using the Club Vita data as the characterising population. Section 10.5.1.2 has highlighted a few time series options to model these trends seen in historical Club Vita data. We consider in more depth:

- VAR with linear trend
- VAR around a constant

By choosing these time series for the time dependent variables in the M7-M5 model, we either assume a continuation of the historical trend when projecting or keeping it static respectively.

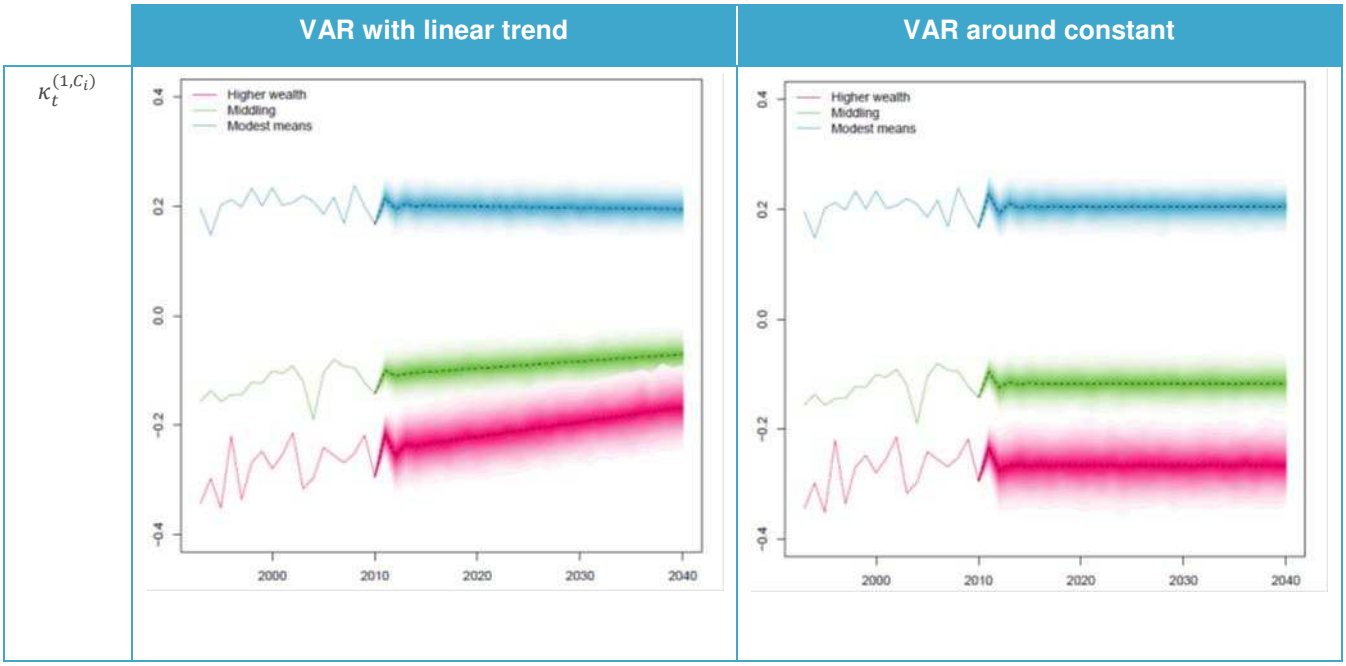


Figure 11.4a: Projections of the fitted parameters for $\kappa_t^{(1,C_i)}$, split by three Club Vita clusters, using both VAR around a constant (left hand side) and VAR with linear trend (right hand side).

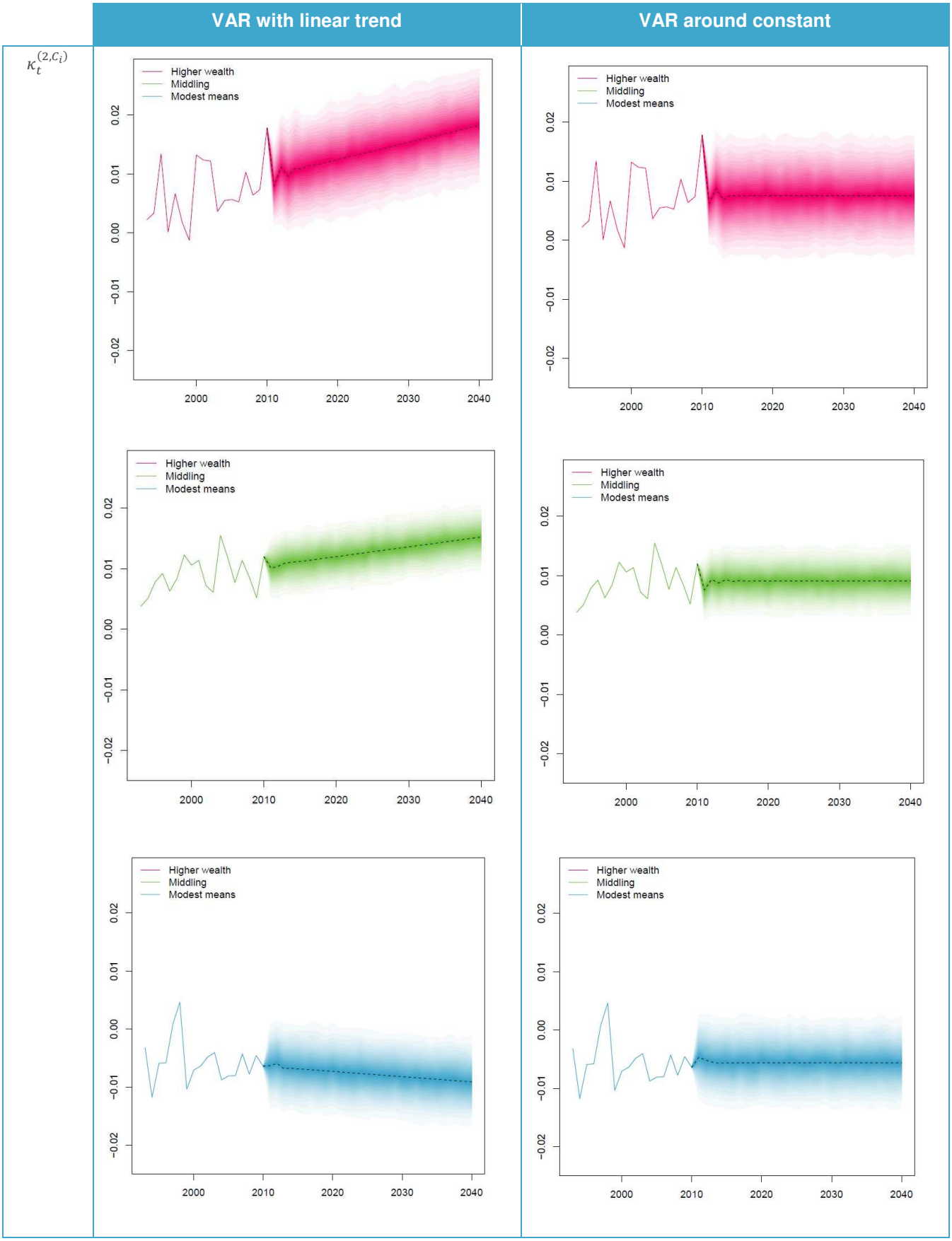


Figure 11.4b: Projections of the fitted parameters for $\kappa_t^{(2,C_i)}$, split by three Club Vita clusters, using both VAR with linear trend (left hand side) and VAR around a constant (right hand side).

As highlighted in section 10.5.1.2, care needs to be taken when using a VAR with trend to model the historical trend in the Club Vita data since the trend in the 'higher wealth' group might be strong enough to overtake the other groups leading to counter-intuitive relative levels of mortality. This is hinted at on the left hand side of Figure 11.4a where the higher end of the funnel for the **higher wealth** group crosses the lower end of the funnel for the **middling** group. On this basis, the time depending variables in the M7-M5 model, $\kappa_t^{(1,C_i)}$ and $\kappa_t^{(2,C_i)}$ will be modelled by VAR with constant in this case study for the Club Vita dataset (See Appendix C for more details of these time series).

11.2.1 Simulations under the VAR with constant

The charts below demonstrate the simulated period life expectancies at age 65 for each cluster in the Club Vita data, in 2010 (left) and 2020 (right).

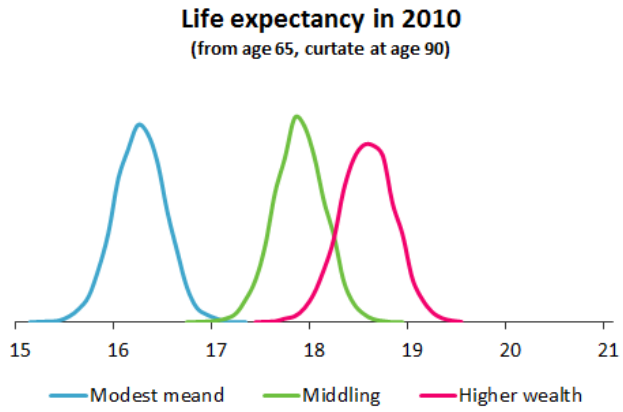


Figure 11.5a: Increase in life expectancy from 2010 to 2020 at age 65.

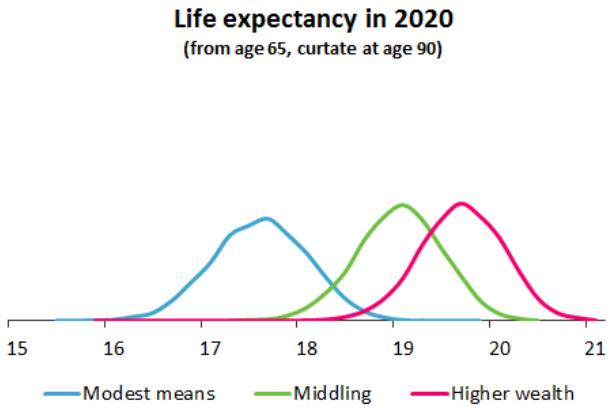


Figure 11.5b: Increase in life expectancy from 2010 to 2020 at age 65.

The density distributions of the simulated life expectancies at age 65 in 2010 (Figure 11.5a) demonstrate a clear separation between the three Club Vita clusters, all in a reasonable order. As in the ONS case, the narrow spread in the distributions reflects parameter uncertainty only due to 2010 being the last calendar year of the historical data fitted to, and no projections have been implemented at this stage[59].

Looking at the simulations in 2020 (Figure 11.5b) we observe a much wider spread of outcomes in the life expectancy distributions compared to 2010 (as in the case when using the ONS data) since we now have ten years of compounded process risk.

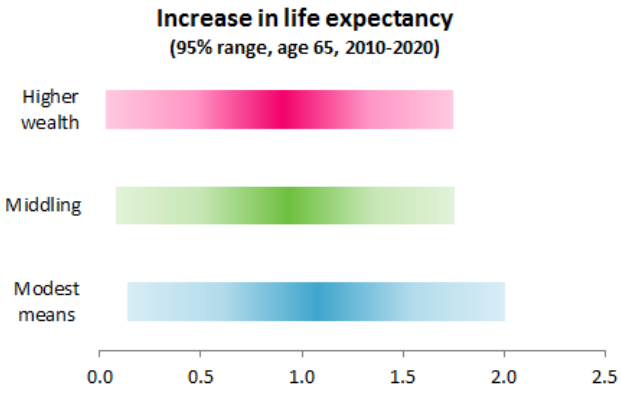


Figure 11.5c: Increase in life expectancy from 2010 to 2020 at age 65.

Despite the overlapping of densities in 2020, it is reassuring that on closer inspection of the projected differences in life expectancies we find a small probability of the **middling** are outliving the **higher wealth** (0.3%) or of the **modest means** outliving the **middling**.

The **modest means** group is also projected to experience slightly faster increases in life expectancy between 2010 and 2020 (Figure 11.5c).

[59] The sampling risk will be included when these simulated rates will be applied to the book

11.3 Hedge effectiveness results

In Sections 11.1 and 11.2 we have seen how different characterisation data result in different assumptions and implications when generating future mortality projections. The following analysis highlights the hedge effectiveness results for a selection of pension schemes when taken through the characterisation approach.

11.3.1 Test schemes

Choosing relevant pension schemes for this case study on the characterisation approach is important since ideally we would like to test the impact of the approach on a wide selection of different schemes. The following schemes were chosen to demonstrate the results of applying the characterisation approach

Scheme	Annual exposure[1]	Exposure period	IMD split Least Middling Most Unknown	Club Vita Wealthy Middling Unhealthy Unknown	Commentary
Large A	28k	1993 2011 2013			• Single scheme • Large enough to do direct modelling • Long history
Large B	28k	1995 2007 1993 2013			• Combined scheme[2] • Large enough to do direct modelling • Medium history
Large C	28k	1997 2006 1993 2013			• Combined scheme[2] • Large enough to do direct modelling • Medium history
Medium	20k	1997 2006 1993 2013			• Single scheme • Borderline for direct modelling • Medium history • Wealthy
Small	12k	1993 2011 2013			• Single scheme • Too small for direct modelling • Long history • Very wealthy

Different (lives) splits

Notes:
1. Exposure in final year of data
2. Combined schemes are generated by pooling data from schemes in very similar industries to create a sufficiently large portfolio for direct modelling.

Figure 11.6: Illustrating the test schemes from the Club Vita dataset

Looking at each scheme in turn:

Large Scheme A

- The scheme is clearly large enough to be modelled directly (with annual exposure of around 28,000) and contains very long back history (1993-2011).

Large Scheme B and C

- Similar characteristics as Scheme A, i.e. large enough for direct modelling but with less back history available.
- Scheme C has been chosen, in particular, for the potential of giving particularly different results for the direct modelling and the characterisation approach since;
  - It is borderline in terms of available history for direct modelling
  - The scheme has considerable number of records where the characterising variables (pension amount or IMD score) are not available (grey segment in the pie charts). This might lead to some unexpected noise in the results for the characterisation approach.

Medium scheme

- Includes around 20,000 annual exposures covering data from 1997 to 2011.
- This scheme has been chosen to show how results of the two modelling approaches (direct modelling and characterisation approach) compare using a scheme which is of borderline size and history length to be modelled directly.

Small scheme

- Relatively small scheme with annual exposure around 12,000 lives.
- This scheme has been chosen to show results of the characterisation approach when applied in typical situations, i.e. when the scheme is too small to be taken through the direct modelling approach.

11.3.2 Results

The table below shows the variance reduction[60] for period survival probabilities[61] age 70 to 90 measured 10 years into the forecasting for the five chosen sample schemes. Each of the columns labelled (3) to (5) in the table shows the hedge effectiveness for the schemes for a different characterising population / time series assumption.

(1) Example scheme	(2) Direct Modelling M7-M5 (VAR with Constant)	(3) Club Vita Characterisation (VAR with Constant)	(4) ONS Characterisation (MRWD)	(5) 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	N/A	75%	70%	79%

Table 11.1: Variance reduction for period survival probabilities age 70 to 90 (measured 10 years from the last year of historical book data)

Whilst only a small selection of schemes, looking at the table we can draw a series of informative observations.

11.3.2.1 Direct modelling vs Club Vita (VAR with constant)

First we start by comparing columns (2) and (3) of the table which show the variance reduction for the sample schemes derived from applying the direct modelling approach (column 2) on the one hand and the characterisation approach (using Club Vita data) on the other (column 3).

The time series assumptions are directly comparable in these two approaches i.e. VAR with constant. Thus, comparisons between these columns help inform us about the level of adjustment that may be needed for residual basis risk not captured by the characterisation approach. We can see that:

- The characterisation approach provides a credible alternative to direct modelling as the results are very similar (Large scheme C aside – see below).

[60] Variance Reduction $= 1 - Var({}_{20}p^{B}_{70,10} - {}_{20}p^{R}_{70,10}) / Var({}_{20}p^{B}_{70,10})$

[61] Notice the different hedge effectiveness statistics ${}_{20}p_{70}$ being used at this stage of the report compared to one used in the direct modelling section. This reflects the run-off approach used for simulating the book under the characterisation approach as described in Figure 10.7.

- For each of the three large schemes the differences in variance reductions within the schemes are modest so that any allowance for residual basis risk is likely to be small.
- The characterisation approach may suggest lower hedge effectiveness than direct modelling (Large Scheme B), and may therefore be more informative about the underlying demographic risk than direct modelling i.e. it is clearly a helpful alternative to direct modelling, even for books large enough to be modelled directly.
- Scheme C exhibits large differences in variance reduction (around 12%). This difference is likely to be a consequence of a combination of shorter history providing less certainty for direct modelling, and more variability in the characterisation results due to the significant amount of unknown records. Consequently this could be viewed as providing an indication of how disparate the results between the two methods could be.
- For the **medium scheme** we observe similar results as in the cases for large schemes A and B. This is reassuring since this scheme is a borderline case of being appropriate for the direct modelling approach.

11.3.2.2 Club Vita (VAR with constant) vs ONS (VAR around tend)

Whilst we saw in section 11.1 that the VAR with trend is unlikely to be an appropriate time series for modelling the ONS dataset it is instructive to consider the impact of changing the dataset without changing the nature of the underlying time series (in terms of bounded or unbounded variability). For this reason we have included the ONS (VAR around trend) results in column (5) of the table[62].

Focussing on columns (3) and (5), it is reassuring to see we observe similar magnitudes of hedge effectiveness when using either the ONS data or the Club Vita data as the characterisation population combined with a consistent choice of time series (i.e. bounded variability).

11.3.2.3 ONS (MRWD) vs ONS (VAR around trend)

By comparing the results in column (4) and (5) we notice how the adoption of a more appropriate time series model (MRWD) for the ONS dataset noticeably changes the hedge effectiveness results.

The difference in variance reduction, around 10%, is comparable to the difference previously observed in Figure 6.10 of section 6 for different choices for the structure of our two population model, i.e. the choice of time series is potentially as important as a model choice decision and so requires careful user engagement. However, the choice of the times series will be specific to the nature of the characterising population and to what views the user will wish to embed in the projections. As such it is less amenable to the kind of structured framework / decision tree we have provided for the model choice.

We can also see how the greater complexity (and therefore wider potential variability in results) in the structure of MRWD time series model, coupled with the use of drift rather than constant trend over time, leads, as we would expect, to a greater residual basis risk than under both direct modelling and the characterisation approach with a VAR time series (columns 3 and 5).

11.3.2.4 Overall perspective on basis risk

Overall, we generally observe a material and meaningful variance reduction (and thus hedge effectiveness) in the table above with broadly similar results across the schemes for those time series approaches which appear reasonable (i.e. we see around a 10% spread between columns 2, 3 & 4).

[62] Note that we have included the ONS (VAR around trend) as column (5) rather than column (4) to keep the most credible results (columns (2)-(4)) together

This is reassuring since it indicates that the characterisation approach is not only a relevant solution for small schemes when hedging basis risk but also a good alternative for those schemes which are large enough to be modelled directly.

12 Practicalities of characterisation approach

At the heart of the characterisation approach is direct modelling of multi-populations. As such the challenges and practicalities described in Section 9 apply equally to the characterisation approach.

There are also a number of additional practical considerations of which the user needs to be aware.

12.1 Characterising variables

When mapping the book population, care is needed to ensure that the mapping is 'consistent' and to handle missing variables.

12.1.1 Consistent variables

Some variables may have a very different meaning in the book population and the characterising population. For example an insurer with a book of individual annuities would not be able to use the pension bands described in section 10 without some modification, since the typical individual retail annuity payment is considerably smaller than the typical DB pension.

12.1.2 Missing or unassignable variables

We will often find that one or more of the variables needed to map the book population onto the characterising population are only available for some individuals. Whilst we would generally expect pension/annuity amount to be available, both of the example characterisations given in section 11 have relied on postcode being available. This is not always the case – it may be absent, or mistyped – or indeed where it is available it may be overseas or outside the region (e.g. England) for which we have a consistent measure of deprivation and so 'unassignable'.

In order to apply the characterisation approach we need a means of assigning these individuals to a characterising subpopulation. One such pragmatic approach, using the known splits within the book population, has been described in Section 10.6. This effectively assumes that (conditional on the known values for any other characterising variables) the variable of concern is 'missing at random'. Generally this will be a reasonable assumption, but does introduce noise and potentially additional basis risk.

It is therefore preferable to reduce the number of individuals for whom this is an issue. Data cleaning can help in this regard (to correct mistyped postcodes etc…). It may also be possible to extend the coverage of the characterising variable – for example by creating an IMD index which is consistent across England, Scotland, Wales and Northern Ireland as per Payne & Abel (2012)[63].

12.2 Projecting men and women

Under the characterisation approach we will have sufficient data volumes for the characterising groups for both men and women. Consequently this removes the need to carry out the unisex modelling described in section 9.2.2. However care will be needed to ensure that the projections for the characterising groups for men and women are coherent, particularly when they have a directly comparable meaning. For example, when the characterising population is the ONS IMD data we would want the resulting simulations for a specific IMD to retain the ordering of male mortality being higher than female mortality.

It is likely, therefore, that, as a minimum, simultaneous modelling of the male and female reference population will be required. This could be achieved for instance with the use of a multivariate time series model for the period indices of men and women (see Li and Hardy (2011)). It may also be necessary to extend the

[63] An application of this approach is illustrated in the joint National Association of Pension Funds (NAPF) and Club Vita publication 'The NAPF Longevity Model' (NAPF, 2014) with the details of the construction of a UK-wide IMD score available in the supporting Technical Appendices (Club Vita, 2014)

multivariate time series modelling of the differences between the book and reference population so that the model captures correlations within gender between characterising groups, and between genders within each characterising group.

12.3 Applying user judgement

As well as needing to apply judgement on the nature of the time series used in modelling the characterising groups, there are other areas where there is potential for user judgement, most notably the correlations assumed between the characterising groups. The year-on-year correlations are usually fitted as a time-invariant parameter to the historical data, i.e. the average level of correlation seen historically. However the user may have a reason to vary these correlations as the simulation evolves, e.g. if they hold the view that future drivers of longevity will apply less equitably across society.

However, care is needed when modifying the correlation matrices in order to ensure that the resulting matrices remain positive semi-definite at each future point in time and so continue to be valid as correlation matrices; and that the projections remain coherent (in terms of the ordering of mortality).

References

Adams (2012) The Canadian Pensioners Mortality Table Historical Trends in Mortality Improvement and a Proposed Projection Model Based on CPP/QPP Data as at December 31 , 2007 *Paper to Canadian Institute of Actuaries* Retrieved from http://www.cia-ica.ca/docs/default-source/2013/213012e.pdf

Ahmadi, S. S., and Li, J. S.-H. (2014). Coherent mortality forecasting with generalized linear models: A modified time-transformation approach. *Insurance: Mathematics and Economics*, 59, 194–221.

Banks, J, Breeze E., Lessof, C. and Nazroo, J. eds. (2006) Retirement, health and relationships of the older population in England. The 2004 English Longitudinal Study of Ageing, *Institute for Fiscal Studies, UCL, NatCen.*

Baxter, S.D. Should projections of mortality improvements be subject to a minimum value? *British Actuarial Journal 13(III),* 375-478

Biatat, V. D., & Currie, I. D. (2010). Joint models for classification and comparison of mortality in different countries. In *Proceedings of 25rd International Workshop on Statistical Modelling, Glasgow* (pp. 89-94).

Booth, H., Hyndman, R. J., Tickle, L., & de Jong, P. (2006). Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, *15*, 289–310.

Börger, M., Fleischer, D., & Kuksin, N. (2013). Modeling the Mortality Trend Under Modern Solvency Regimes. ASTIN Bulletin, 44(01), 1–38.

Brockwell, Peter J., and Richard A. Davis, eds. *Introduction to time series and forecasting*. Vol. 1. Taylor & Francis, 2002

Butt, Z., & Haberman, S. (2009). llc: a collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms. *Working paper.*

Cairns, A. J. G., Blake, D., & Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 687-718

Cairns, A. J. G., Blake, D., Dowd, K., & Coughlan, G. D. (2011a). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41, 29–59.

Cairns, A. J. G., Dowd, K., Blake, D., & Coughlan, G. D. (2013). Longevity hedge effectiveness: a decomposition. *Quantitative Finance 14*(2) 217-235

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011b). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, *48*(3), 355–367.

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

Carter, L. R., & Lee, R. D. (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, 8(3), 393–411.

Chatfield, C. (2013) The analysis of time series: an introduction. *CRC press*

Club Vita (2014) NAPF Longevity model – Supporting technical appendices. Available from: http://www.clubvita.co.uk/SiteCollectionDocuments/NAPF%20Technical%20Report.pdf

Continuous Mortality Investigation (2007). Working Paper 25 - Stochastic projection methodologies: Lee-Carter model features, example results and implications.

Continuous Mortality Investigation (2011). Working Paper 53 – An initial investigation into rates of mortality improvement for pensioners of self-administered pension schemes

Continuous Mortality Investigation (2014) Working Paper 73 - Analysis of the mortality experience of pensioners of self-administered pension schemes for the period 2005-2012.

Currie, I. (2009). Piggyback Forecasts of Mortality. Adjusting for bias in mortality forecasts. *Presentation to Joining Forces on Mortality and Longevity, 22 October 2009, Edinburgh.*

Currie, I. D. (2014). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal 0*(0), 1-28

Danesi I., Haberman S., & Millossovich P. (2014) Forecasting Mortality in Subpopulations Using Lee-Carter Type Models: A Comparison. *Working Paper.*

Debón, A., Montes, F., & Martínez-Ruiz, F. (2011). Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. *European Actuarial Journal, 1*(2), 291-308.

Delwarde, A., Denuit, M., Guillen, M., & Vidiella, A. (2006). Application of the Poisson log-bilinear projection model to the G5 mortality experience. *Belgian Actuarial Bulletin, 6*(1), 54-68.

Dixon, A., Le Grand, J., Henderson, J., Murray, R. & Poteliakhoff, E. (2007) Is the British National Health Service equitable? The evidence on socioeconomic differences in utilization *Journal of Health Services Research and Policy, 12,* 104-109.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., & Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal, 15*(2), 334-356

Evandrou, E. & Falkingham, J. (2002) Smoking behaviour and socio-economic class: a cohort analysis, 1974 to 1998 Health Statistics Quarterly, 14, 30-38

Forfar, D.O., McCutcheon, J.J., & Wilkie, A.D., 1988. On graduation by mathematical formula. Journal of the Institute of Actuaries 115, 1–135.

Gavrilov, L.A. & Gavrilova, N.S. (1991) The biology of life span: a quantitative approach *Harwood Academic, 1991*

Goddard, M. & Smith, P. (2001) Equity of access to health care services, *Social Science & Medicine, 53*, 1149-1162

Greenland, S. (2001). Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. International Journal of Epidemiology, 30(6), 1343–50.

Haberman, S., & Renshaw, A. E. (2009). On age-period-cohort parametric mortality rate projections. Insurance: Mathematics and Economics, 45(2), 255–270.

Haberman, S., & Renshaw, A. (2011). A comparative study of parametric mortality projection models. Insurance: Mathematics and Economics, 48(1), 35–55.

Hatzopoulos, P., Haberman, S. (2013). Common mortality modeling and coherent forecasts. An empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, *52*(2), 320-337.

Hoffmann, R. (2005). Do socioeconomic mortality differences decrease with rising age? *Demographic Research*, *13*, 35–62.

Human Mortality Database. (2013). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). *www.mortality.org or www.humanmortality.de*.

Hunt, A., & Blake, D. (2014a). A General Procedure for Constructing Mortality Models. North American Actuarial Journal, 18(1), 116–138.

Hunt, A., & Blake, D. (2014b). Identifiability in Age/Period/Cohort Mortality Models. *Working paper.*

Hunt, A., & Villegas, A. M. (2014). Robustness and convergence in the Lee-Carter model with cohorts. *Under Review*

Hymans Robertson LLP (2014) Buy-outs, buy-ins and longevity hedging, Q2 2014 *Available from: http://www.hymans.co.uk/media/591601/140905-managing-pension-scheme-risk-q2-2014.pdf*

Jarner, S. F., & Kryger, E. M. (2011). Modelling adult mortality in small populations: The Saint Model. *ASTIN Bulletin*, *41*(2), 377–418.

Jarner, S. F., & Møller, T. (2013). A partial internal model for longevity risk. *Scandinavian Actuarial Journal*, (ahead-of-print), 1-31.

Kaufman, L. & Rousseeuw, P.J. (2005) Finding Groups in Data: An Introduction to cluster Analysis *Wiley Series in Probability and Mathematical Statistics*

Kleinow, T. (2013). A Common Age Effect Model for the Mortality of Multiple Populations *In Longevity 9 Conference, Beijing*

Koissi, M., Shapiro, a, & Hognas, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, *38*(1), 1–20.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659–671.

Li, J. S.-H., & Hardy, M. R. (2011). Measuring basis risk in longevity hedges. North American Actuarial Journal, 15(2), 177–200.

Li, N. & Lee R (2005) Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography* 42.3 (2005): 575-594.

Li, J. S., Zhou, R., & Hardy, M. R. (2014). A Step-by-Step Guide to Building Two-Population Stochastic Mortality Models, 1–35.

London School of Economics (LSE) for the Equality and Human Rights Commission: Socio-economic inequalities in older people's access to and use of public services. Retrieved from http://justageing.equalityhumanrights.com/

Lütkepohl, Helmut. *New introduction to multiple time series analysis*. Springer, 2007.

Madrigal, A.M., Matthews, F.E., Patel, D.D., Gaches, A.T. & Baxter, S.D. (2012) What Longevity Predictors Should Be Allowed For When Valuing Pension Scheme Liabilities *British Actuarial Journal,* 16(1), 1-38

McCullagh, P., & Nelder, J. (1989). Generalized Linear Models (Second Edi.). London: Chapman & Hall.

Morris, S., Sutton, M. and Gravelle, H. (2005) Inequity and inequality in the use of health care in England: An empirical investigation *Social Science & Medicine, 60*, 1251-66.

Mosher, J & Sagoo, P (2011) Longevity: Parallels with the past *The Actuary, October 2011 Edition*

NAPF (2014) The NAPF Longevity Model – November 2014
Available from: http://www.napf.co.uk/PolicyandResearch/DocumentLibrary/0414_longevity_model_nov14.aspx

Olshansky, S.J., Passaro, D.J., Hershow, R.C., Layden, J., Carnes, B.A., Brody, J., Hayflick, L., Butler, R.N., Allison, D.B. & Ludwig, D.S. (2005) A potential decline in life expectancy in the United States in the 21st Century. *New England Journal of Medicine* **352**, 11 1138-1145

ONS (2011) A report on the 2009 General Lifestyle Survey
Available from http://www.ons.gov.uk/ons/rel/ghs/general-lifestyle-survey/2009-report/index.html

ONS (2014) Life Expectancy at Birth and at Age 65 by Local Areas in the United Kingdom, 2006-08 to 2010-12.
Available at http://www.ons.gov.uk/ons/dcp171778_360047.pdf

Payne, R.A., & Abel, G.A. (2012) UK indices of multiple deprivation - a way to make comparisons across constituent countries easier *Health Statistics Quarterly 53*

Plat, R. (2009a). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3), 393–404.

Plat, R. (2009b). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics*, *45*(1), 123–132.

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3), 556–570.

Renshaw, A. E., & Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee–Carter modelling. *Insurance: Mathematics and Economics*, *42*(2), 797–816.

Rose & Pevalin (2005) The National Statistics Socio-economic Classification: Origins, Development and Use. *Institute for Social and Economic Research, University of Essex*

Russolillo, M., Giordano, G. & Haberman, S. (2011). Extending the Lee–Carter model: a three-way decomposition. *Scandinavian Actuarial Journal*, *2011*(2), 96-117.

Salhi, Y. & Loisel, S. (2013) Longevity Basis risk modelling: a co-integration based approach. *Working paper.*

Scholes, S., Bajekal, M., Love, H., Hawkins, N., Raine, R., O'Flaherty, M., & Capewell, S. (2012). Persistent socioeconomic inequalities in cardiovascular risk factors in England over 1994-2008: a time-trend analysis of repeated cross-sectional data. *BMC Public Health*, *12*(1), 129.

Unal, B., Critchley, J.A. & Capewell, S (2004) Explaining the decline in Coronary Heart Disease mortality in England & Wales between 1981 and 2000 *Circulation* **109**, 1101-1107

Villegas, A., Bajekal, M. and Haberman, S. (2014). Modelling mortality by cause of death and socio-economic stratification: an analysis of mortality differentials in England. *Presentation at international mortality and longevity symposium 2014, 17 September 2014, Birmingham.*

Villegas, A., & Haberman, S. (2014). On the Modeling and Forecasting of Socioeconomic Mortality Differentials: An Application to Deprivation and Mortality in England. *North American Actuarial Journal*, *18*(1), 168–193.

Yang, S. S., & Wang, C.-W. (2013). Pricing and securitization of multi-country longevity risk with mortality dependence. Insurance: Mathematics and Economics, 52(2), 157–169.

Wan, C., Bertschi, L., & Yang, Y. (2013). Coherent mortality forecasting for small populations : an application to Swiss mortality data. In AFIR/ERM-PBSS-Life Colloquium, Lyon.

WHO (2003) Adherence to long-term therapies: Evidence for action *Published by the World Health* Organisation Retrieved from: http://www.who.int/chp/knowledge/publications/adherence_full_report.pdf

Zhou, R., Li, J. S.-H., & Tan, K. S. (2013). Pricing Standardized Mortality Securitizations: A Two-Population Model With Transitory Jump Effects. *Journal of Risk and Insurance*, 80(3), 733–774.

Zhou, R., Wang, Y., Kaufhold, K., Li, J. S.-H., & Tan, K. S. (2014). Modeling Period Effects in Multi-Population Mortality Models: Applications to Solvency II. North American Actuarial Journal, 18(1), 150–167.

Appendix A: Club Vita data

A1 Introduction to Club Vita

Club Vita was founded in 2007 with a primary focus on helping pension schemes understand one of their major risks - longevity. In return for submitting membership data, participating pension schemes receive analysis of the longevity of their members. The main aim of this analysis is to help these schemes understand the unique characteristics at individual scheme level and appreciate the consequences of how they are likely to change over time.

The participating schemes span a wide range of sizes including some of the largest DB schemes, which are comparable in (monetary) size to the annuity books of larger insurers. The varying sizes of the schemes enable us to test both the direct modelling approach and characterisation approach.

A2 Club Vita dataset

The Club Vita database (VitaBank) is a pool of data of individual pension scheme member records, submitted by the member schemes. This database (as at September 2014) consists of nearly 6 million member records; including:

- **2.1 million** pensioners and widow(er)s;
- **3.7 million** members below the retirement age; and
- **0.7 million** deaths.

The records received include personal, but non-sensitive information from pension scheme administration, i.e. information relevant to predicting longevity, such as date of birth, postcode, pension, salary and retirement health are collected.

A number of checks are carried out on the data received to ensure it is correct and reliable, and where possible corrections are made. Where a member record has a predictor which our checks suggest is unreliable it is excluded from analysis of the impact of that predictor. We also check for concentrations of unreliable records within schemes, and biases in exclusions between living and deceased records and limit a scheme's inclusion in our analysis accordingly.

By using this cleaned data Club Vita is able to analyse and understand the impact of longevity on affluence, gender, occupation, lifestyle and other predictive rating factors, and how these impacts have been changing over time.

A3 Data extract used in this analysis

Club Vita collects data annually for each of its subscribers with these data feeds spread over the calendar year. As such it is regularly refreshed with the latest longevity data.

For the purposes of our analysis into longevity basis risk we have needed to have a static dataset. Accordingly we have focussed on an extract of the database as at January 2013 and have used this throughout our analysis.

The growing number of schemes participating in Club Vita mean that this extract contains fewer records then are currently held in the Club Vita dataset: with approximately 12.9 million life years of exposure and 0.4 million deaths.

A2.1 Exposed to risk

The chart below shows the pattern of (pensioner and dependant) exposures over time for men (blue bars) and women (pink bars) within the data analysed in this report.

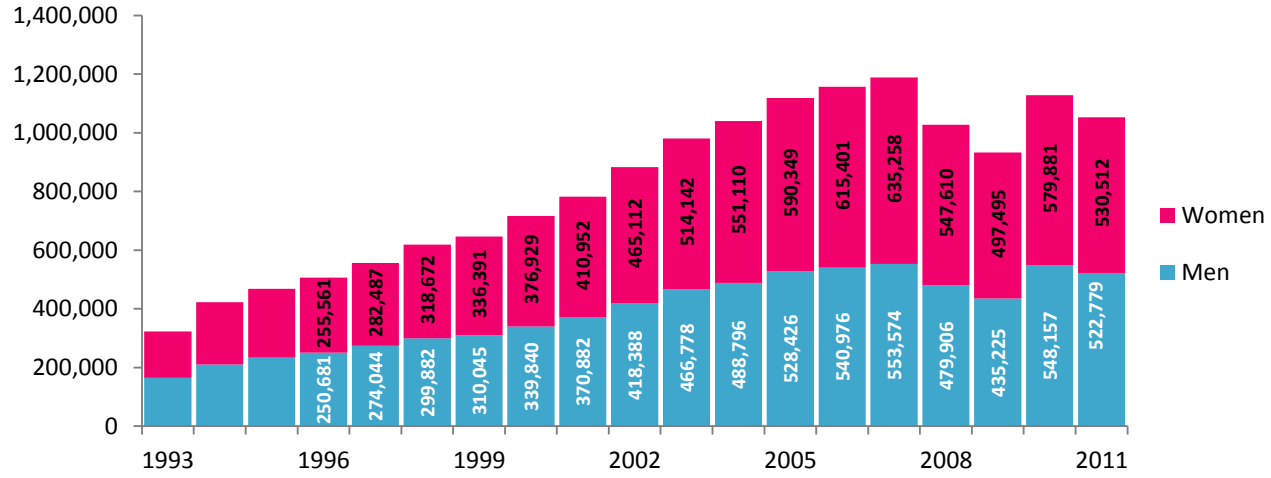


Figure A1: Exposed to risk by calendar year for the Club Vita data analysed in this report

We can see how:

- The exposures increase over time reflecting
  - schemes within the Club having reliable data starting at different points in time due to historical administration practices; and
  - the maturation of pension schemes leading to larger numbers of pensioners
- There is a drop in exposures around 2008 and 2009. This reflects a number of schemes which, having participated in a free initial pilot, did not continue to subscribe to Club Vita.
- The exposures for 2011 are low compared to earlier years – owing to this being a partial year of exposure for many schemes due to when they had last submitted data prior to the point of data extract

A4.2 Deaths

The chart below shows the number of deaths each year in the dataset we have used. This follows a similar pattern to the exposed to risk.

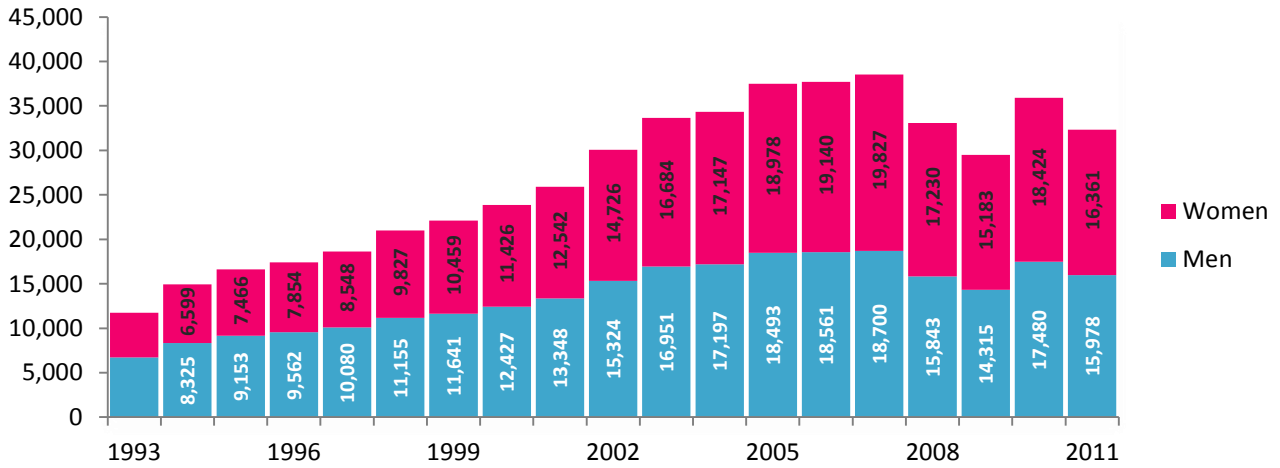


Figure A2: Deaths by calendar year for the Club Vita data analysed in this report

A4 Key rating factors in VitaBank

By virtue of collecting information at the individual level, VitaBank contains a wide range of rating factors potentially relevant to both baseline mortality and improvements therein. These rating factors include gender, retirement health, pensioner type, lifestyle (geo-demographics), affluence, age and occupation (manual and non-manual)[64]. We briefly discuss some of the rating factors used in our analysis below.

A4.1 Lifestyle

The lifestyle variable is created by using a third party service provider of geo-demographic data (ACORN) which maps UK postcodes onto a demographic profile specific to the full postcode.

Within Club Vita we have condensed these 57 different ACORN categories using statistical clustering methods into 7 different lifestyle categories which are predictive of material differences in longevity.

A4.2 Deprivation (IMD)

An alternative postcode based socio-economic measure is deprivation. Deprivation encompasses of the following domains; financial, health, education, service or crime scores. These domains are used by the national statistics agencies to construct Index of Multiple Deprivation (IMD) scores within each of England, Scotland, Wales and Northern Ireland.

We have used IMD as a rating factor for improvement patterns as it is easily accessible and a well-known measure for deprivation; and can if needed be easily restated onto a consistent basis across all four countries using the technique described in Payne & Abel (2012). We have focussed on quintiles of IMD for England within our analysis. The chart below shows the distribution of our data between quintiles ranging from the least deprived (Q1) to the most deprived (Q5).

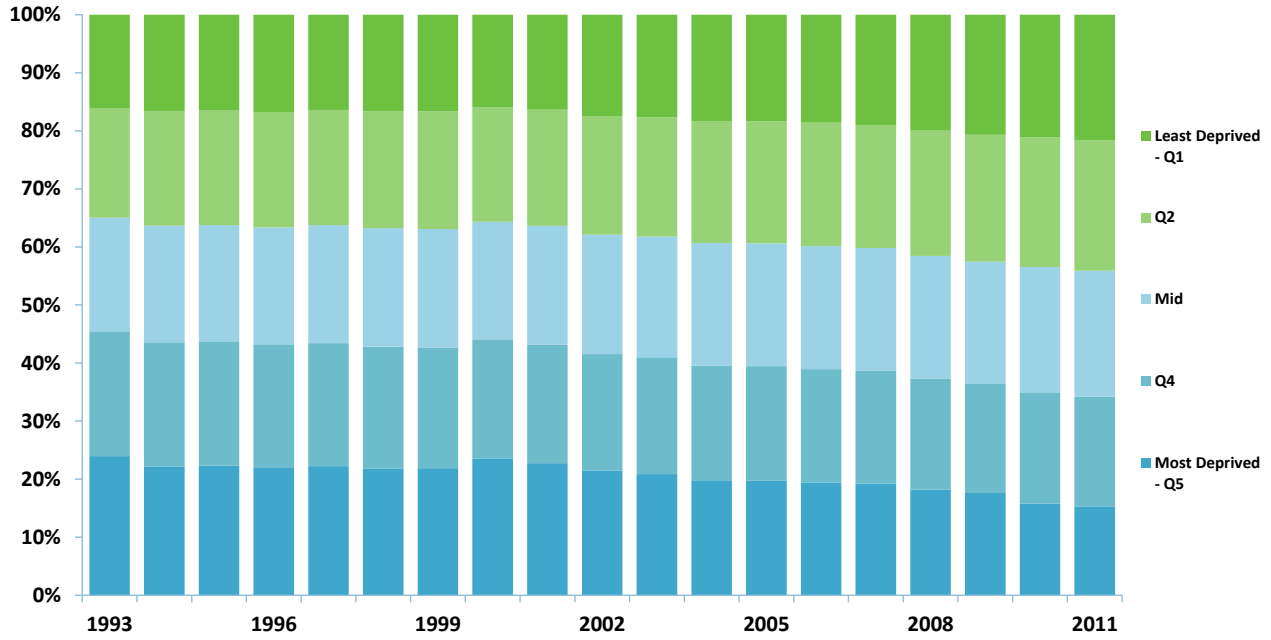


Figure A3: Split of Club Vita data by deprivation quintile

We can see how over time there is a growing propensity towards the least deprived quintiles (Q1 and Q2) and declining proportions in the most deprived quintiles (Q4 and Q5). This will reflect, at least in part, the survivorship bias whereby within any cohort of retirees, a greater proportion of those in the least deprived quintiles will survive to the most recent calendar years.

[64] See Madrigal et al (2012) for more detail on how Club Vita have determined the key ratings factors for mortality levels

For both deprivation and lifestyle we have relied on using an individual's postcode at date of death when assigning a value for historical death records. This implicitly assumes that where gentrification of an area has occurred over time this is would not change the broad grouping (lifestyle or deprivation quintile) that would apply (see e.g., Appendix D in Lu et. al (2012)). For lifestyle groupings which use the full six digit postcode this is likely to be more of an issue than for the deprivation score focussed on here, which are measured at the broader local super output area; these are less sensitive to (but not immune to) the gentrification of particular streets / neighbourhoods.

A4.3 Pension and salary

The Club Vita data contains two measures of affluence: pension and last known salary (usually the last salary at retirement).

Pension can be a poor proxy to affluence as it depends not only on earnings but length of service in the pension scheme – a modest pension could arise from long service on low pay, or very short service on high pay. However, whilst salary is a better measure of affluence, pension will generally be available, whereas salary may be harder to extract from some pension scheme records.

To allow for inflation both pension and salary are revalued from their as at date to a common date (1 January 2010) in line with RPI. For deceased pensioners the revaluations occur at a proportion of RPI (below 100%) for broad consistency with the pension increases paid historically to surviving pensioners which will typically be a mix of full RPI, limited price inflation and nil increases.

Appendix B: Landscape of two-population models

The table below sets out the key features of each the models considered in sections 5 & 6.

Model	References	Reference population $\boldsymbol{logit(q^R_{xt})}$	Book population $logit(q^B_{xt}) - logit(q^R_{xt})$	Key features
Stratified Lee-Carter	• Butt & Haberman (2009) • Debon et all (2011)	$\alpha^R_x + \beta^R_x \kappa^R_t$	α^B	• Age independent level differences[65] • Equal improvements
Piggyback model	• Currie (2009)	$\alpha^R_x + \beta^R_x \kappa^R_t$	$a^B + b^B x$	• Linear in age level differences • Equal improvements
Common-Factor	• Carter & Lee (1992) • Li & Lee (2005) • Li & Hardy (2011)	$\alpha^R_x + \beta^R_x \kappa^R_t$	α^B_x	• Age specific level differences • Equal improvements
Three-Way Lee-Carter	• Russolillo et al.(2011)	$\alpha^R_x + \beta^R_x \kappa^R_t$	$\alpha^B_x + \lambda^B \beta^R_x \kappa^R_t$	• Age specific level differences • Trivial (perfectly correlated) improvements
Joint-κ	• Carter and Lee (1992) • Li and Hardy (2011) • Wilmoth and Valkonen(2001) • Delwarde et al. (2006)	$\alpha^R_x + \beta^R_x \kappa^R_t$	$\alpha^B_x + \beta^B_x \kappa^R_t$	• Age specific level differences • Trivial (perfectly correlated) improvements
Common-age effect	• Kleinow (2013)	$\alpha^R_x + \sum_j \beta^{(j,R)}_x \kappa^{(j,R)}_t$	$\alpha^B_x + \sum_j \beta^{(j,R)}_x \kappa^{(j,B)}_t$	• Age specific level differences • Non perfectly correlated improvements • Same age response to change in period terms

[65] Level and improvement differences here refer to the logit level and improvements.

Model	References	Reference population $\boldsymbol{logit(q_{xt}^R)}$	Book population $logit(q_{xt}^B) - logit(q_{xt}^R)$	Key features
Lee-Carter + VAR/VECM	• Zhou et al. (2013)	Cannot be formulated in relative terms		• Requires book and reference histories with same length
Cointegrated Lee-Carter	• Carter and Lee (1992) • Li and Hardy (2011) • Yang and Wang (2013)	Cannot be formulated in relative terms		• Requires book and reference histories with same length
Augmented-Common Factor	• Li and Lee (2005) • Li and Hardy (2011) • Hyndman et al. (2013)	$\alpha_x^R + \beta_x^R \kappa_t^R$	$\alpha_x^B + \beta_x^B \kappa_t^B$	• Age specific level differences • Non perfectly correlated improvements
Relative Lee-Carter + cohort	• Villegas and Haberman (2013)	$\alpha_x^R + \beta_x^R \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \beta_x^B \kappa_t^B$	• Age specific level differences • Non perfectly correlated improvements • Allows for cohort effect
Gravity Model - Two population APC	• Dowd et al. (2011)	$\alpha_x^R + \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \kappa_t^B + \gamma_{t-x}^B$	• Separate age, period and cohort effect differences additively
Bayesian Two Population - Two population APC	• Cairns et al. (2011)	$\alpha_x^R + \kappa_t^R + \gamma_{t-x}^R$	$\alpha_x^B + \kappa_t^B + \gamma_{t-x}^B$	• Accounts for short book histories • Relatively complex to implement
Two Population CBD – M5	• Li et al (2014))	$\kappa_t^{(1,R)} + (x - \overline{x})\kappa_t^{(2,R)}$	$\kappa_t^{(1,B)} + (x - \overline{x})\kappa_t^{(2,B)}$	• Linear in age level differences • Period indices give level and slope • Non perfectly correlated improvements

Model	References	Reference population $logit(q^R_{xt})$	Book population $logit(q^B_{xt}) - logit(q^R_{xt})$	Key features
Two Population M6	• Li et al (2014)	$\kappa_t^{(1,R)} + (x-\overline{x})\kappa_t^{(2,R)} + \gamma^R_{t-x}$	$\kappa_t^{(1,B)} + (x-\overline{x})\kappa_t^{(2,B)} + \gamma^B_{t-x}$	• Linear in age level differences • Two period indices give level and slope • Non perfectly correlated improvements • Allows for cohort differences
Two Population M7	• Li et al (2014)	$\kappa_t^{(1,R)} + (x-\overline{x})\kappa_t^{(2,R)} + ((x-\overline{x})^2 - \sigma_x^2)\kappa_t^{(3,R)} + \gamma^R_{t-x}$	$\kappa_t^{(1,B)} + (x-\overline{x})\kappa_t^{(2,B)} + ((x-\overline{x})^2 - \sigma_x^2)\kappa_t^{(3,B)} + \gamma^B_{t-x}$	• Quadratic in age level differences • Period indices give level, slope and curvature • Non perfectly correlated improvements • Allows for cohort differences
Saint model	• Jarner and Kryger (2011) • Jarner and Moller (2013))	Frailty based model	$\kappa_t^{(1,B)} + (x-\overline{x})\kappa_t^{(2,B)} + ((x-\overline{x})^2 - \sigma_x^2)\kappa_t^{(3,B)}$	• Can be reformulated as a model with parametric age term
Co-integration Approach	• Salhi and Loisel (2013)	Cannot be formulated in relative terms		• Sequence of age-by-age models • Complex to implement • Can hardly accommodate a cohort effect
Plat Relative Model	• Plat (2009)	$\kappa_t^{(1,R)} + (x-\overline{x})\kappa_t^{(2,R)}$	$\frac{100-x}{100-\bar{x}}\kappa_t^{(1,B)}$	• Can be reformulated as a model with parametric age term
Relative P-Splines	• Biatat and Currie (2010)	Complex p-splines formula		• Complex to implement • Forecasting may be problematic • inclusion of cohort-effect is non-trivial

Model	References	Reference population $logit(q_{xt}^R)$	Book population $logit(q_{xt}^B) - logit(q_{xt}^R)$	Key features
Plat+Lee-Carter	Wan et al. (2013)	$\alpha_x^R + \kappa_t^{(1,R)} + (x - \bar{x})\kappa_t^{(2,R)} + \gamma_{t-x}^R,$	$\alpha_x^B + \sum_{j=1}^{M} \beta_x^{(j,B)} \kappa_t^{(j,B)}$	• Non perfectly correlated improvements
Multipopulation GLM	• Hatzopoulos and Haberman (2013) • Ahmadi and Lee (2014)	Generalised linear modelling formulations		• Complex to implement • Not simple to understand

Appendix C: Overview of time series

This appendix contains a brief and concise overview of the series used in the project. Extensive introduction to time series, their properties and forecasting procedures can be found in Brockwell and Davis (2002), Chatfield (2013) and Lütkepohl (2007).

C1 Introducing time series

A time series is made of a sequence of observations

$$X_1, \dots X_t, \dots$$

where the index is interpreted as time. A time series is *stationary* (weakly or second-order) if the first two moments exist, $E(X_t)$ is constant and $Cov(X_{t+h}, X_t)$ only depends on h.

C2 Stationary time series

A first example of stationary time series is the autoregressive of order 1, or AR(1), which is the solution of the equation

$$X_t = \phi_0 + \phi_1 X_{t-1} + \xi_t$$

where ξ_t is a white noise, that is a sequence of uncorrelated, zero mean, variables. The process is (asymptotically) stationary if $|\phi_1| < 1$. This type of dynamics implies that the long term behaviour of the sequence approaches a stable, time independent distribution.

C3 Non-stationary time series

A typical example of a non-stationary process is one including a trend. Such a process can be modelled using an *integrated* process, defined through the first order difference $\Delta X_t = X_t - X_{t-1}$. A process is autoregressive of order 1, integrated of order 1, or ARI(1,1), if ΔX_t is an AR(1), that is

$$\Delta X_t = \phi_0 + \phi_1 \Delta X_{t-1} + \xi_t.$$

The special case $\phi_1 = 0$ is known as random walk with drift, or RWD. Its dynamic can be written as

$$X_t = \phi_0 + X_{t-1} + \xi_t.$$

A straightforward calculation shows that the mean and variance of X_t grow linearly with time.

C4 Multivariate time series

Coming to multivariate time series, we consider a sequence

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_t, \dots$$

where each term is a vector composed by d observations,

$$\boldsymbol{X}_t = \begin{bmatrix} X_{t1} \\ \vdots \\ X_{td} \end{bmatrix}.$$

Such a process is stationary if the vector $E(X_t)$ is constant and the matrix $Cov(X_{t+h}, X_t)$ depends on h only. A vector autoregressive model of order 1, or VAR(1), satisfies the equation

$$\boldsymbol{X}_t = \Phi_0 + \Phi_1 \boldsymbol{X}_{t-1} + \boldsymbol{\xi}_t \ ,$$

where Φ_0 is a column vector, Φ_1 is a square matrix and $\boldsymbol{\xi}_t$ is a multivariate white noise, that is a sequence of uncorrelated zero-mean variables. The correlation between the d observations in $\boldsymbol{X}_t$ comes from the

dependence on all variables at time $t-1$ and the covariance matrix of $\boldsymbol{\xi}_t$. Stationarity is guaranteed if the modulus of the eigenvalues of A are smaller than 1.

C5 Incorporating a trend in multivariate time series

A process incorporating a trend can be modelled explicitly using a VAR(1) around a deterministic trend or implicitly with a multivariate random walk with drift (MRWD). A VAR(1) around a deterministic trend is specified by the following equation:

$$\boldsymbol{X}_t = \Phi_{00} + \Phi_{01} t + \Phi_1 \boldsymbol{X}_{t-1} + \boldsymbol{\xi}_t \ ,$$

where Φ_{00} and Φ_{01} are vectors of intercepts and slope parameters. A MRWD satisfies the following equation:

$$\boldsymbol{X}_t = \Phi_0 + \boldsymbol{X}_{t-1} + \boldsymbol{\xi}_t \ .$$

Although both models allow for linear trends in the mean, they differ in the variability that is permitted around this trend, with the VAR(1) having bounded variability and the MRWD having unbounded variability.

Appendix D: Characterising groups for the Club Vita dataset

The Club Vita DB annuitant dataset enables us to group individuals by a wide range of characteristics which could be indicative of differences in future longevity trends and thus demographic risk. In Section 2.2.1 we saw that – of those characteristics which are widely available within pension schemes / BPA books – combining pension income and postcode based deprivation measures provided the most parsimonious fit to historical experience. In this Appendix we describe in more detail the steps taken to create the characterising groups across pension and IMD for the Club Vita Dataset. These steps (and underlying thought processes) could be repeated on alternative multi-dimension characterising population datasets available to the user.

D1 Overview of clustering approach

We create our characterising groups via the following steps:

1 Segment the spectrum of values each variable can take into discrete groups, and so in turn the characterising population into 'cells' (D2)

2 Identify a distance metric which measures the level of dissimilarity between these cells striking a balance between the competing principles given in section 10.4 (D3)

3 Use statistical techniques (specifically partitioning about medoids and fuzzy analysis) to cluster these cells into our desired number of groups (D4)

4 Interpret the results of the clustering and consider whether it is appropriate to adjust the allocation of cells to ensure groups are both interpretable and credible in size (D5)

D2 Segmenting the characterising population into cells

Both deprivation and affluence are measured on a continuous spectrum. One possible approach to identify cells is to use optimisation techniques to identify the splits of pension and deprivation which optimise the resulting clusters in terms of their fit to historical trends. This would be computationally more complex and time-consuming, and liable to the fallacy of self-prophecy.

Given the desire to form sensible groups, and the acceptance that the past is at best a guide to the future, we take a more pragmatic approach of splitting each variable. Both the deprivation and the IMD spectrum have been split broadly into quintiles, thus forming 25 cells which is reassuringly an order of magnitude larger than the number of characterising groups we are aiming for (3), yet sufficiently few to avoid excessive noise when measuring mortality rates and improvements for these cells.

For deprivation we have used the publicly available IMD quintiles, accepting that the distribution of pensioners may be skewed a little away from the most deprived quintiles and so will not form perfect quintiles within the Club Vita dataset.

For pension we have identified the Club Vita quintiles within the data, and then rounded these to avoid spuriously precise cut-offs between groups. The resulting bands are (in 1 January 2010 monetary terms); below £2,000 p.a., £2,000 - £3,000 p.a., £3,000 - £5,000 p.a., £5,000 - £10,000 p.a., and over £10,000 p.a.

The resulting cells can be visualised as per the table to the right. The percentages within each cell refer to the proportion of the Club Vita dataset which lies in each cell. There is an element of concentration of the data along the leading diagonal (top left – bottom right) of the table consistent with an element of correlation between affluence (pension) and socio-economics (deprivation).

Pen/IMD	Q5 (Most deprived)	Q4	Q3 (Mid)	Q2	Q1 (Least deprived)	Total
<2k	6%	6%	6%	5%	4%	27%
2-3k	3%	3%	3%	3%	2%	14%
3-5k	5%	4%	4%	4%	3%	20%
5-10k	4%	4%	5%	5%	4%	22%
10k+	1%	2%	3%	5%	6%	17%
Total	19%	19%	21%	22%	19%	100%

D3 Dissimilarity matrix

In order to be able to apply standard statistical clustering techniques we need a measure of 'distance' between the cells in order to group cells which are 'closest' together. Principles 2-5 from section 10.4 provide us with three dimensions across which to measure the distance:

1 **Characteristics:** The similarity of characterising cells in terms of the underlying variables which define the cell e.g. pension and deprivation. *(Our interpretability principle)*

2 **Mortality levels:** The similarity of cells in terms of the levels of mortality. *(Our mortality levels principle)*

3 **Mortality improvements:** The similarity or otherwise of cells in terms of observed mortality improvements (*Our principles of grouping similar improvements, but separating clear differences.)*

Formally this resulting 'distance' between cells is expressed in the form of a **dissimilarity matrix**. If there are n characterising cells then the dissimilarity matrix is an $n \times n$ lower triangular matrix d where d_{ij} is the 'distance' (dissimilarity) between cluster cell i and clustering cell j.

A wide range of distance metrics exist and could be used to measure the dissimilarity between cells. In choosing a metric we need to be sensitive to having a mix of ordinal variables (e.g. IMD quintile or pension band) and nominal variables (mortality levels and improvements). We follow the approach suggested in Kaufman & Rousseeuw (2005) whereby each of the variables is measured using an interval scaled approach. This involves converting each dimension to a numerical quantity and then measuring the distance between two cells as the absolute value of the difference in these quantities, divided by the maximum absolute value the difference takes.

Taking each dimension in turn:

Characteristic dimension: We rank the characteristics in their natural order (high to low deprivation; order of increasing pension band) and use the ranks as our numerical value. We apply equal weight to the ranks and then interval scale.

Mortality levels: As a nominal value we can interval scale this by taking the difference in mortality levels. However, different cells may have difference age distributions, and as such care is needed not to confound differences in mortality levels arising due to differences in age distribution. We broadly control for this by calculating the mortality levels for three different age bands (65-74, 75-84, 85-94) and giving each equal weight in the interval scaling. In each case we use the mortality levels for 2008-2010.

Mortality improvements: By applying linear regression to the crude q_{xt} for each characterising group we obtain a smoothed average annual improvement rate (as the gradient of the linear regression) to which we can apply interval scaling. As improvements are liable to be different at different ages we also need to control for possible differences in average age by calculating this for different age bands (65-74, 75-84, 85-94). We have annualised the improvements over the whole period for which we have reliable data, i.e. 1993-2010. If the user were to be worried about changing patterns over time, for example certain cells seeing slower improvements earlier in the period and faster improvements later in the period, it would be possible to refine this by creating separate smoothed improvement rates for different parts of the period 1993-2010. Care is needed not to create too many improvements dimensions as this could lead to the clustering being driven by statistical noise rather than genuine differences.

We then take a weighted average of the dissimilarities in each dimension to give the overall dissimilarity.

Formally we therefore have:

$$d_{ij} = \omega_{characteristics}\, d_{ij}^{characteristics} + \omega_{levels}\, d_{ij}^{levels} + \omega_{improvements}\, d_{ij}^{improvements}$$

$$\omega_{characteristics} + \omega_{levels} + \omega_{improvements} = 1$$

with the individual distances given by:

$$d_{ij}^{characteristics} = \frac{1}{2}\left(\frac{|Rank\ cell\ i\ for\ pension - Rank\ cell\ j\ for\ pension|}{4} + \frac{|Rank\ cell\ i\ for\ IMD - Rank\ cell\ j\ for\ IMD|}{4}\right)$$

$$d_{ij}^{levels} = \frac{1}{3}\left(\frac{\left|q_{65-74,2008-2010}^{cell\ i} - q_{65-74,2008-2010}^{cell\ j}\right|}{max.\ abs\ diff\ in\ mortality\ ages\ 65-74} + \frac{\left|q_{75-84,2008-2010}^{cell\ i} - q_{75-84,2008-2010}^{cell\ j}\right|}{max.\ abs\ diff\ in\ mortality\ ages\ 75-84} + \frac{\left|q_{85-94,2008-2010}^{cell\ i} - q_{85-94,2008-2010}^{cell\ j}\right|}{max.abs\ diff\ in\ mortality\ ages\ 85-94}\right)$$

$$d_{ij}^{improvements} = \frac{1}{3}\left(\frac{\left|\Delta q_{65-74,2008-2010}^{cell\ i} - \Delta q_{65-74,2008-2010}^{cell\ j}\right|}{max.\ abs\ diff\ in\ improvements\ ages\ 65-74} + \frac{\left|\Delta q_{75-84,2008-2010}^{cell\ i} - \Delta q_{75-84,2008-2010}^{cell\ j}\right|}{max.\ abs\ diff\ in\ improvements\ ages\ 75-84} + \frac{\left|\Delta q_{85-94,2008-2010}^{cell\ i} - \Delta q_{85-94,2008-2010}^{cell\ j}\right|}{max.\ abs\ diff\ in\ improvements\ ages\ 85-94}\right)$$

Here the Δq notation is used to signify the average annual improvement rate derived from the linear regression, and *max. abs diff* refers to the maximum absolute difference

What weights to use?

Having defined our measures of dissimilarity our final decision is what weights we apply to the different dimensions, i.e. the values for $\omega_{characteristics}$, ω_{levels}, and $\omega_{improvements}$. From a purely theoretical perspective we are able to make a number of observations:

- The greater the weight given to $\omega_{characteristics}$ the more liable the method is to group contiguous blocks within the clusters.
- However, the construction of $d_{ij}^{characteristics}$ favours forming 'crosses' within the grid of characterising cells. This is because it considers the cells marked B below to be closer to A (0.125) than the cell marked C is to A (0.25)

	B	C
	A	B

- In mortality terms it is likely that cell C is closer to cell A than either of the cells marked B is, as the reduction in pension in moving from cell A to cell C may be compensated by the reduction in deprivation. This suggests we should provide slightly more weight to the mortality levels than the characteristics, as the mortality levels are both a principle in their own right, but also aid in creating interpretable clusters.
- We need to have some weight to improvements ($\omega_{improvements}$) as this is our variable of interest – however putting too much weight on improvements risks clustering very disparate cells purely due to the noise in observed improvements. This suggests we may wish to limit $\omega_{improvements}$ to around 50%.

Whilst we have not sought to optimise the exact weightings used, we have tried a variety of combinations and found that setting $\omega_{characteristics} = 0.2$, $\omega_{levels} = 0.3$, and $\omega_{improvements} = 0.5$ enabled us to identify interpretable clustering groups with meaningful differences in mortality and improvements.

D4 Cluster cells

Using the dissimilarity matrix we can cluster cells which have the lowest dissimilarity (i.e. most alike in terms of a blend of characteristics, mortality levels and mortality improvements). A host of statistical techniques exist for doing this. We have looked at the results under two approaches[66]:

- **Partitioning about medoids (PAM):** Partitioning about medoids selects a single cell per desired cluster to be *representative* of that cluster. The remaining cells are then clustered with whichever of these representative cells they have the smallest dissimilarity. By varying the initial choice of representative cells the algorithm seeks to minimise the aggregate dissimilarity, i.e. the sum of the dissimilarities between each cell and the representative cell with which it is clustered.
- **Fuzzy analysis:** Fuzzy analysis seeks to minimise a (weighted) sum of the dissimilarities between the cells within each cluster, However, rather than allocating each cell to a cluster, it instead considers that each cell could be split between clusters i.e. belong, in part, to one or more clusters. The proportions in which each cell is split between the clusters is optimised.

Both methods are readily available to users through statistical software such as R. One advantage of the fuzzy analysis approach is that it avoids condensing each group down to a single representative cell at each stage and so retains more of the underlying information at each stage of the cluster. However, this also makes it technically a more complex method to understand.

Another advantage of the fuzzy analysis approach is that it produces a form of 'probability' associated with each cell belonging to each of the three groups, i.e. the optimised proportions. This enables the user to apply judgement where some reallocation of cells to groups might improve the credibility of size of the group or aid the interpretation of the resulting groups.

Applying these two methods using a dissimilarity matrix with 20% weight to characteristics, 30% weight to mortality levels and 50% weight to mortality improvements suggests the following groupings, where for the fuzzy analysis we have shown the groupings implied by allocating each cell to the group with which it has the greatest associated probability:

[66] For more information on these, and other clustering methods, see Kaufman & Rousseeuw (2005)

Partitioning about medoids

Pen/IMD	Q5	Q4	Q3	Q2	Q1
<2k			X		
2-3k		X			
3-5k					
5-10k				X	
10k+					

Cells marked in darker shade and with an X are the medoids for each group.

Fuzzy analysis

Pen/IMD	Q5	Q4	Q3	Q2	Q1
<2k	88%	78%	64%	65%	59%
2-3k	87%	69%	35%	73%	41%
3-5k	83%	61%	39%	57%	65%
5-10k	78%	44%	49%	50%	47%
10k+	48%	72%	55%	66%	60%

Percentages in individual cells refer to the proportion of the cell which would be allocated to the coloured group. This is the highest 'probability' for each cell. Each cell will also have 'probabilities' of being in the other groups produced by the analysis.

D5 Interpreting the results

Reassuringly the results of the two methods are broadly consistent with just 5 cells allocated to different groups under the two approaches. The main ambiguities over allocation of cells relate to the mid deprivation (Q3) for pensions up to £10k, and the low deprivation / low pension combinations shown in pink under partitioning about medoids (PAM) and green under fuzzy analysis.

Under the PAM partitioning we can see how we can readily get groups with a natural interpretation if we reallocate the £2-3k pension / Q3 IMD cell to the green group. We would then have a low income / high deprivation group, a wealthy group, and a middling group. In this case though the wealthy group is likely to be very large and so the characterising groups may fail to adequately discriminate between individuals.

In contrast, the fuzzy analysis provides us with 'probabilities' as to which group each cell lies in. Looking at the ambiguous cells we have:

Cell		"Probability" belongs to			Notes
Pension	IMD	Blue	Green	Pink	
£2-3k	Q1	19%	41%	40%	Pink under PAM, Green under fuzzy
£3-5k	Q1	7%	64%	28%	Pink under PAM, Green under fuzzy
£3-5k	Q2	13%	57%	30%	Pink under PAM, Green under fuzzy
<£2k	Q3	18%	64%	18%	Green under both approaches, but to aid interpretation would want to be blue
£2-3k	Q3	35%	34%	32%	If <2k, Q3 cannot be blue then this would ideally be green to get a natural interpretation to the groups
£3-5k	Q3	39%	33%	28%	Green under PAM, Blue under fuzzy
£5-10k	Q3	11%	49%	40%	Pink under PAM, Green under fuzzy

We can see how:

- we can be confident that Q1 and Q2 cells in the £3-5k row are better placed in the green group rather than the pink group as suggested by the PAM approach *(second & third rows of table)*
- the £2-3k Q1 combination which is pink under the PAM method and green under fuzzy analysis is a borderline case – however given the certainty of the green groupings in the £3-5k row for Q1 and Q2 we propose that it be kept as green to avoid an isolated cell *(first row of table)*
- within the Q3 IMD column
  - the below £2k cell is clearly green (64%) and so to get groupings with natural meanings the rest of the column should be green or pink *(fourth row)*
  - the £2-3k cell can reasonably reallocated from the blue group to the green group *(fifth row)*
  - the grouping of the £3-5k Q3 cell is ambiguous and it would be reasonable to include this in the green middling group rather than the blue group *(sixth row)*
  - the £5-10k Q3 group could reasonably be in the pink (wealthy) group *(seventh row)*

This leads us to three groups which could be described as the **'modest means'**, **'middling'** and the **'higher wealth'**.

Pen/IMD	Q5 (most)	Q4	Q3 (mid)	Q2	Q1 (least)
<5k					
5-10k					
10k+					

Appendix E: Generation of synthetic data

In this appendix we present a procedure for generating, based on a reference dataset, synthetic mortality datasets which have a given exposure size with a given distribution of this exposure across population subgroups.

Assume that we have a reference dataset containing observed number of deaths D_{xtg} in year t for people age x in subgroup g with matching central exposures E_{xtg} and matching death rates μ_{xtg}=D_{xtg}/E_{xtg}.

Let C'_t be the target total exposure for year t in the synthetic dataset and $(w'_{tg_1}, \dots, w'_{tg_m})$ be a vector of weights adding to one which represents the splitting of this exposure among the subgroups.

The synthetic central exposures E'_{xtg} in year t for people age x in subgroup g are obtained as

$$E'_{xtg} = C'_t \frac{\sum_g E_{xtg}}{\sum_x \sum_g E_{xtg}} w'_{tg} = C'_t \frac{E_{xt}}{E_t} w'_{tg}$$

where $E_{xt} = \sum_g E_{xtg}$ are the total exposed to risk at age x in year t across all groups and $E_t = \sum_x \sum_g E_{xtg}$ are the total exposed to risk in year t across all groups and ages. Hence the exposure for the reference dataset is being used to obtain the split by age for a particular year and group. The corresponding synthetic number of deaths D'_{xtg} is generated by drawing a random sample from a Poisson distribution with mean $E'_{xtg}\mu_{xtg}$.

Appendix F: Original call for proposals (abridged)

This appendix contains sections 1 to 4 of the original call for proposals issued by the LBRWG and has been reproduced with the LBRWG's permission.

Project to develop a method of assessing basis risk for longevity transactions

F1 Summary

This document is an invitation to tender for a research project for the Longevity Basis Risk Working Group (LBRWG).

The aim of the project is:

- **to develop a readily-applicable methodology for quantifying the basis risk arising from the use of population-based mortality indices for managing the longevity risk inherent in specific blocks of pension benefits or annuitant liabilities.**

The methodology will be statistically rigorous and practical: it will use data likely to be available in respect of the population and the block of business being hedged.

The LBRWG has received a commitment to fund the project from the Institute and Faculty of Actuaries (IFoA) and the Life and Longevity Markets Association (LLMA), subject to receipt of a satisfactory proposal and to achievement of interim project targets.

We believe this project will offer the successful party an opportunity to use statistical knowledge and/or original research to produce a solution to a real industry problem. If the project were successful and facilitated the transfer of longevity risk between market participants, the work would be ground-breaking and very high-profile. We would expect that the methodology would use the indices published by the LLMA but be applicable in any territory world-wide subject to the availability of appropriate data.

We expect the project to last between 12-18m from the time the project is awarded; further details of the timeline are set down below. However, credible proposals that could be completed in a shorter time frame would be considered.

We are seeking proposals from actuarial consultancies and academic institutions. Responses to the tender should be received by Monday 15th April 2013.

F2 Background to the project

The LLMA began publishing indices linked to population mortality statistics in March 2012 with the goal of facilitating the hedging of longevity risk for pension funds and annuity books. The launch of the LLMA indices was an important milestone towards a longevity market where risk management can be carried out through transactions that are linked to standardised population-level data. Index-based hedges have considerable potential to provide effective risk and capital management for all holders of longevity risk.

In addition to the mortality indices, the LLMA has also produced a significant body of work around possible derivative transactions that could reference mortality indices and offer 'standardised' longevity risk management tools (see www.LLMA.org.uk/Library).

However the building blocks described above have not proved sufficient to develop a 'liquid market' in longevity and have not led to transactions based on these standardized measures. Indeed, both are underutilised relative to more traditional longevity transactions that occur in the market. Some institutions currently use risk management tools linked to indices – the concept is proven. Even so, we believe that a major obstacle to widespread use of longevity risk management tools that reference population-based mortality indices is the difficulty in quantifying, and hence managing, **longevity basis risk**.

There are two major considerations for longevity basis risk:
- The need to understand the nature of the risk and its impact in different circumstances, and
- The need to account for the basis risk underlying the transaction in reported results.

In December 2011 the LLMA and IFoA formed the LBRWG […]. Its remit is straightforward: to investigate how to provide a market-friendly means of analysing longevity basis risk.

Having carefully considered the matter, we have concluded that the task is beyond the scope of the working group by itself. The challenge is technically complex and time-consuming. Further research, or considerable work to synthesise existing research, is required before a solution can be developed. So we require the assistance of either a consultancy firm or an academic/research institution to perform that research.

A short summary of the work done to date by the working group is outlined in […].

F3 A description of Longevity Basis Risk

Longevity basis risk is the potential mismatch between the behaviour of a longevity hedge and the portfolio of pensioners or annuitants being hedged, in cases where the hedging transaction's cash flows are determined by reference to a mortality index and not directly linked to the actual pool of lives.

There are three major sources of basis risk between the pension fund/annuity book risk to be hedged and the value of the hedging tools employed to reduce that risk. These are:

- **Demographic risk**: the difference between μ_1 and μ_2, the underlying forces of mortality for the reference portfolio and the pension fund/annuity book, respectively, due to demographic or socio-economic differences. This difference may comprise two elements: the initial (current) level of mortality and the rates of future improvement.
- **Sampling risk**: the difference in the population sizes (exposures) and varying levels of annuity amounts, because any sub-population is a random sample of the large population, so the observed mortality rates in the two populations will not be the same, except by chance.
- **Structural risk** due to the payoff structure of the hedge. We could for example use a portfolio of S-Forward derivatives and compare how the value of that portfolio behaves versus the original liabilities being hedged (see the LLMA website for a description of S-Forward hedges). The pay out of the hedge is unlikely to exactly match the liabilities being hedged.

These three sources of basis risk all contribute to a longevity hedge being a less-than-perfect match to the portfolio being hedged. We believe that demographic risk and sampling risk are most usefully analysed through stochastic projections of mortality rates. Structural risk can be analysed relatively simply after the other two, because structural risk can be quantified in a straightforward fashion once scenarios of mortality rates have been projected for the different populations under consideration. Such quantification involves calculating the value of the hedge instrument under every scenario of mortality and then looking at the expected value of the result, either in isolation or relative to the pension or annuity portfolio value using a relevant metric. Therefore

defining and optimising a hedge portfolio is a separate exercise from trying to estimate the relationship between the progression of mortality behaviours between μ_1 and μ_2.

F4 The proposal

Throughout the project, the goal will be an outcome that is *practically applicable* to analysing basis risk arising from standard information available to a regular market participant. Original academic research may be required, but only in so far as it leads towards that goal.

Our proposal is for an overall project delivered in two phases:

Phase 1 would be the demonstration of the feasibility of a methodology for determining the relationship between μ_1 and μ_2 in the future.

Deliverables for Phase 1 would be:

- Details of relevant **background research**, including:
  - a review of evidence of different mortality improvement rates among different subgroups (e.g. by socioeconomic group, affluence or location) to inform underlying assumptions and structure of relationship between μ_1 and μ_2 in projection methodology;
  - a critical review of existing models for the structure of the relationship between μ_1 and μ_2 in projection methodologies, in light of above review of evidence;
- Detailed specification of a **proposed methodology**, to include a general description and a detailed technical/statistical analysis;
- Analysis of the **limitations of the methodology** and a description of any alternative methodologies that may have been considered with an explanation of why the proposed methodology best achieves the aims of the project;
- A clear specification of the **work to be completed**, and the **anticipated outputs** from that work, in Phase 2.

Funding for Phase 2 would be dependent on satisfactory completion of Phase 1, to be determined by the LBRWG and the sponsoring organisations. The LBRWG and sponsoring organisations would need to be satisfied that the aims of the project remained realistically achievable.

Phase 2 would be the practical application of the Phase 1 work to demonstrate the use of the initial research in practice.

Deliverables for Phase 2 would be:

- Definition of **metrics covered by the proposed model** and a demonstration of how the outputs from the methodology can be used for those metrics;
- **Application of the model** on practical, realistic, illustrative examples based on the data reasonably available to potential users;
- Demonstration of how the outputs from the model can be presented as a **robust quantification of basis risk** to third parties such as regulators.

Institute
and Faculty
of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 · Exchange Crescent · 7 Conference Square · Edinburgh · EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax: +44 (0) 131 240 1313

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Hong Kong

2202 Tower Two · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418 · Fax: +11 (0) 852 2147 2497

www.actuaries.org.uk