City University London

Doctoral Thesis

Mortality Modelling and Longevity Risk Management

Author:
Andrew Hunt

Supervisor:
Prof. David Blake

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Faculty of Actuarial Science and Insurance
Cass Business School

16th September 2015

Dedicated to the grey wanderer, Oðin Allfather, the grim god of the gallows and patron and protector of this humble scrivener.

Contents

List of Figures

List of Tables

Acknowledgements

Many people have contributed directly or indirectly to the work presented in this thesis. I would especially like to thank:

- Bent Nielsen for discussions around the topic of identifiability in mortality models and maximal invariants, and suggestions regarding Chapters 2, 3 and 4;
- Michele Bergamelli for help with the cointegration techniques used in Chapters 7 and 8, and for general econometric wisdom;
- Frank van Berkum for conversations regarding trend changes and the problems with cohort parameters discussed in Chapter 6;
- Robert Cowell for help with the Bayesian approach for modelling cohort parameters in Chapters 6 and 12;
- Pietro Millossovich, Matthias Börger, Ana Debon, Frank van Berkum and Daniel Harrison FIA for reviewing and making detailed suggestions regarding various chapters;
- participants at the following conferences:
  - the Eighth, Ninth and Tenth International Longevity Conferences in Waterloo, Canada (in September 2012), Beijing, China (in September 2013) and Santiago, Chile (in September 2014), respectively;
  - Perspectives on Actuarial Risks in Talks of Young Researchers in Ascona, Switzerland in January 2013;
  - the 17^{th} International Congress on Insurance: Mathematics and Economics in Copenhagen, Denmark, in July 2013;
  - the 49^{th} Actuarial Research Conference in Santa Barbara, USA in July 2014; and
  - the Society of Actuaries Longevity Seminar in Chicago, USA in February 2015;
- the anonymous referees for Chapters 5 and 8, for their helpful and constructive criticism of those papers prior to publication;
- the Continuous Mortality Investigation for making available the data for the CMI Assured Lives and the Self-Administered Pension Schemes study, used in Chapters 7, 9 and 10; and

- Steven Haberman for suggestions regarding various chapters and general support and good judgement; and

In addition, I would also like to make a few more personal thank yous. First, I would like to thank the examiners of this thesis, Andrew Cairns and Pietro Millossovich for taking the time to read it and give their considered thoughts and criticisms. They have improved the thesis greatly, even without formal corrections, and I am grateful for your insights.

Second, I want to thank my friends and family for keeping me sane, not dismissing the notion of studying for a PhD out of hand and occasionally pretending to be interested in what I was doing. It has been very much appreciated.

In particular, I would like to single out Andrés Villegas for discussions, suggestions and insights beyond number. Without your help, this thesis would be vastly poorer, and without a good friend to bounce ideas off of, the past few years would have been far more lonely. Thank you enormously.

Finally, I would like to thank my supervisor, David Blake. There have been plenty of frustrations along the way, and plenty of things that we still disagree on (whether 'data' is singular or plural springs to mind), but your relentless focus on the big picture and clarity of exposition has been vitally important. If a reader of the following chapters ever finds a part that they think is well explained, or has an example which is clear and helpful, then this is almost certainly your doing. Thank you immensely for all that you have done over the past years and I look forward to continue to collaborate with you in future.

Declaration of Authorship

I, Andrew HUNT, declare that this thesis titled, 'Mortality Modelling and Longevity Risk Management' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.
- I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only simple copies made for study purpose, subject to the normal conditions of acknowledgement.

Signed:

Date:

16th September 2015

CITY UNIVERSITY LONDON

Abstract

Faculty of Actuarial Science and Insurance
Cass Business School

Doctor of Philosophy

Mortality Modelling and Longevity Risk Management

by Andrew HUNT

The 20^{th} century has witnessed some of the largest and most widespread gains in human longevity ever witnessed, which show no sign of slowing down during the early years of the 21^{st} century. The risk of further, higher than anticipated improvements in life expectancy - known as longevity risk - is now a major and growing field of study. This thesis investigates a number of theoretical and practical problems within the field of longevity risk relating to the structure and identifiability issues within many of the most common models used to study mortality rates, the construction of new mortality models, the projection of these models into the future, the impact of differences in the level and evolution of mortality rates in different populations (such as pension schemes) and the market-consistent valuation and measurement of risk in longevity-linked liabilities and securities.

"I advise you to go on living solely to enrage those who are paying your annuities. It is the only pleasure I have left."

Voltaire

"Do do meddle in the affairs of actuaries, for we know when you will die."

Unknown

Chapter 1

Introduction

The 20^{th} century has witnessed some of the largest and most widespread gains in human longevity ever witnessed, which show no sign of slowing down during the early years of the 21^{st} century. Whilst this is overwhelmingly a sign of human progress, the extended period people now expect to spend in retirement has profound financial consequences for those providing pensions and retirement annuities - governments, life assurance companies and pension schemes. Therefore, this risk of further, higher than anticipated improvements in life expectancy – known as longevity risk – is now a major and growing field of study.

Prior to starting my research, I worked as a qualified pensions actuary in the UK. As part of this, I was involved in advising companies and trustees on the options regarding de-risking pension schemes and was seconded to assist with the modelling of the first multi-billion pound longevity swap deal. I have, therefore, seen first-hand that the tools available to quantify and manage longevity risk in pension schemes and annuity books were inadequate to the task. The standard projections of mortality rates used often contained arbitrary and unrealistic assumptions, such as a tailing-off of the rate of improvement, which has been often predicted but has yet to be observed. They were also usually based on national populations with no indication of how the experience of a specific sub-population would be different. Most importantly however, they were deterministic, so were not able to give an indication of the uncertainty due to longevity risk in the liabilities.

As a result, I was often unable to satisfactorily answer many of the questions regarding longevity risk I was asked during the course of my work. These were primarily practical in nature and involved the quantification of longevity risk and its financial implications.

The topics covered in my research have, therefore, attempted to shed light on some of these practical issues.

However, as a result of the dissertation for my MRes (the PhD-level training programme which comprises the first year of the doctoral programme, prior to starting research in earnest), I became increasingly aware that many of the tools developed in academia were also inadequate to providing answers to these questions. Specifically, I found that simple models, such as the benchmark Lee-Carter and Cairns-Blake-Dowd models, were unable to capture the observed behaviour of mortality rates in the historical data and, therefore, would underestimate the potential longevity risk in future. However, more complicated models suffered from a lack of robustness when estimating parameters and complicated identifiability issues within the models. Therefore, my research also needed to enhance the understanding of the issues which limited the practicality of more complicated models and to improve the range of models used to predict mortality rates, before investigating the more practical issues I had encountered in my work.

To achieve these aims for my research, my thesis comprises of four broad parts, each containing chapters which are linked thematically:

- Part I - Structure, Identifiability and Construction of Age/Period/Cohort Mortality Models
  - Chapter 2 - Structure and Classification of Mortality Models
  - Chapter 3 - Identifiability in Age/Period Mortality Models
  - Chapter 4 - Identifiability in Age/Period/Cohort Mortality Models
  - Chapter 5 - A General Procedure for Constructing Mortality Models
- Part II - Projection of Mortality Rates for Single or Multiple Populations
  - Chapter 6 - Consistent Mortality Projections Allowing for Trend Changes and Cohort Effects
  - Chapter 7 - Identifiability, Cointegration and the Gravity Model
  - Chaprer 8 - Modelling Longevity Bonds: The Swiss Re Kortis Bond
- Part III - Modelling Mortality for Pension Schemes
  - Chapter 9 - Basis Risk and Pension Schemes: A Relative Modelling Approach
  - Chapter 10 - Transferring Risk in Pension Schemes via Bespoke Longevity Swaps

- Part IV - Forward Mortality Models
  - Chapter 11 - Forward Mortality Rates in Discrete Time I: Calibration and Securities Pricing
  - Chapter 12 - Forward Mortality Rates in Discrete Time II: Longevity Risk Measurement and Management

These parts form a unified whole and there exist numerous connections between chapters in different parts of the thesis. For instance, all of the models used are from the class of age/period/cohort (APC) mortality models and, therefore, the qualitative understanding of this class developed in Chapter 2 is fundamental to all of the other chapters in this thesis.

During my MRes dissertation, I encountered problems with using the Plat (2009a) model and, especially, the estimation of the cohort parameters within it. In part, I found this was because the model was not fully identified, namely that I needed to apply an additional identifiability constraint on the quadratic trend in the cohort parameters in order to obtain a unique set of parameters when fitting the model to data. This need for an additional identifiability constraint, which was not mentioned in Plat (2009a), made me think more generally about identifiability issues in APC mortality models - both in terms of why they are present and how we can ensure that a model is fully identified. The result of this analysis developed into Chapters 3 and 4, especially as a result of discussing the subject with Bent Nielsen who drew my attention to the impact of identifiability on projections. Furthermore, I had attempted to extend the Plat (2009a) model to younger ages in my MRes, in order to obtain more estimates of more recent cohort parameters. However, my attempts to do so resulted in models which lacked robustness and had terms added in an ad hoc fashion, since I lacked a procedure for extending the model based on the evidence of the data. Overcoming this challenge resulted in the "general procedure" of Chapter 5, which was, itself, only possible once the identifiability issues in more complicated APC models was understood. Thus, the work in Part I followed directly from the issues I encountered during my MRes, but laid the foundation for the subsequent parts of my thesis.

Only once the fundamental structure of APC mortality models was understood and a method for constructing complicated but robust mortality models was devised could I begin on the more practical aspects of modelling longevity risk. This started with the methods used to project mortality rates in national populations, which is discussed in in Part II. These chapters start by looking at a single population and then move on to look

at two population modelling. Much of this work also came from a desire to overcome the modelling issues I had encountered in the MRes, where I had been forced to use to ad hoc "fixes" in order to model the Kortis bond. For example, the development of the Bayesian approach for modelling cohort parameters in Chapter 6 arose from finding that cohort parameters existing on the threshold between being estimated and being projected were poorly estimated and changed dramatically if a different range of the data was chosen, which had large implications for my results. In addition, during my MRes, I has experienced issues with using the gravity model of Dowd et al. (2011b) to project period parameters in a "coherent" fashion. The analysis of these led directly to the discussion of identifiability and cointegration in mortality models in Chapter 7. Putting these together, therefore, Chapter 8 represents an investigation of the same practical issues I had looked at in my MRes dissertation, but armed with the substantially more sophisticated tools I needed to overcome the problems I encountered previously.

This work, however, focused on mortality rates in national populations. My background as a pensions actuary had made me aware that many studies of mortality rates in national populations had limited applicability for the far more data constrained situation faced by a pension scheme. It is this context which informs the work performed in Part III. One of the key questions for many pension scheme actuaries is, assuming we have good models for the projection of mortality in a large national population, how can we quantify the differences between what is observed nationally and the mortality rates in a relatively small pension scheme. Although there have been previous academic studies on this subject, in my opinion most of them were limited by using data for a far larger sub-population than would be typical of a pension scheme (e.g., the CMI Assured Lives dataset). From my work, I knew that the CMI had published data from the Self-Administered Pension Schemes study. In addition to being far more relevant for the investigation of pension scheme mortality rates, this dataset is also available for a far more limited range of years than the datasets typically used in previous studies. It was, therefore, well suited to my purposes. Chapter 9 investigates this dataset using a "relative" modelling approach, and attempts to use this to quantify the potential for "basis risk" (i.e., differences in the evolution of mortality rates in the reference and sub-populations). Chapter 10 then uses this analysis to try to model a stylised pension scheme and so give insights into the value-for-money of bespoke longevity swaps. This is an issue which is of great practical importance, but where I felt very little academic work had been done.

The final part, Part IV, of my research focused on the questions of the measurement and management of longevity risk. One question I was asked by investment professionals

when I was working was what the value at risk of longevity risk was. On considering this issue, I realised that many existing stochastic models were unable to answer the question. This is because the majority of the change in the value of any liability or security linked to longevity relates to changes in expectations of mortality rates beyond the valuation date, rather than changes in the observed rates themselves. To answer this question required the use of forward mortality rates models. However, those which existed were extensions of the Heath-Jarrow-Morton framework for interest rates, which were not designed for mortality rates (and so unable to capture many of the observable features of mortality rates such as cohort effects) and operated in continuous time (which is not compatible with the majority of actuarial valuation techniques in use in practice). Chapters 11 and 12 attempt to revolve this by developing a new technique in discrete time based on expectations of the force of mortality from APC models, and use it for various risk management problems.

This is a long thesis. I make no apologies for this, since I think it attempts to tackle a number of important questions of great practical relevance. Although it is constructed as a series of stand-alone papers, the purpose of this introduction is to illustrate the links between the different chapters and show that the thesis is a unified whole, motivated by a desire to develop and use enhanced modelling tools to understand longevity risk.

However, I am aware that a great deal of further work needs to be done in respect of the modelling of longevity risk. For example, I believe there are interesting extensions to the work in this thesis, such as developing the general procedure in Chapter 5 to allow for heterogeneity in the underlying data, exogenous causal variable such as smoking prevalence or economic factors, or using it for other demographic phenomena (such as fertility rates). I would also like to extend the forward mortality framework in Part IV to value longevity options, incorporate multi-population mortality projections in Chapter 8 and the relative model in Chapter 9 into the framework to allow for basis risk in valuation, and allow for "recalibration risk" (discussed in Chapter 12). I also believe that identifiability in multi-population models, and especially its impact on achieving coherent mortality projections, is a topic worthy of study beyond the relatively brief treatment in Chapter 7.

In summary, my thesis attempts to answer a number of practical questions on the subject to longevity risk, and in doing so has resulted in a better understanding of the framework of the age/period/cohort models and methods for constructing new models. However, there is much which still remains to be done, and plenty of other areas of research which

may prove fruitful in future.

Below is a diagrammatic representation of how the various chapters of this thesis depend upon each other, along with brief abstracts, presentation histories and acknowledgements for each part and chapter, .

1.1 Structure, Identifiability and Construction of Age/Period/Cohort Mortality Models

Much of the analysis of the historical evolution of mortality rates is made using models which decompose mortality rates across the dimensions of age, period and cohort (or year of birth). This includes many of the most widely used mortality models, such as the Lee-Carter, Cairns-Blake-Dowd and classic APC models. However, APC mortality models are not fully identified, which can lead to problems with estimating the parameters within them robustly, and require arbitrary identifiability constraints to be imposed when fitting them to data, which can bias any projections from the model. Part I of the thesis reviews the fundamental structure of APC mortality models, discusses the identifiability issues within them and proposes a "general procedure" for constructing new mortality models which give a superior fit to the historical data.

1.1.1 Structure and Classification of Mortality Models

I am grateful to Andrés Villegas, Steven Haberman, Bent Nielsen and Ana Debón for their detailed comments regarding this chapter.

This chapter provides a holistic analysis of models which examine the structure of mortality rates across the dimensions of age, period and cohort and examines their similarities and differences. Specifically, it investigates the structure of APC mortality models, introduces a classification scheme for existing models and lists the key principles a model user should consider when constructing a new model in this class. This analysis is mainly qualitative in nature and discusses the motivation for many of the subjective judgements made subsequently in the course of the thesis. In addition, since the models used in the remainder of this thesis come from this class, a firm understanding of the APC structure is vital for the development of more sophisticated mortality models and underpins much of the following work

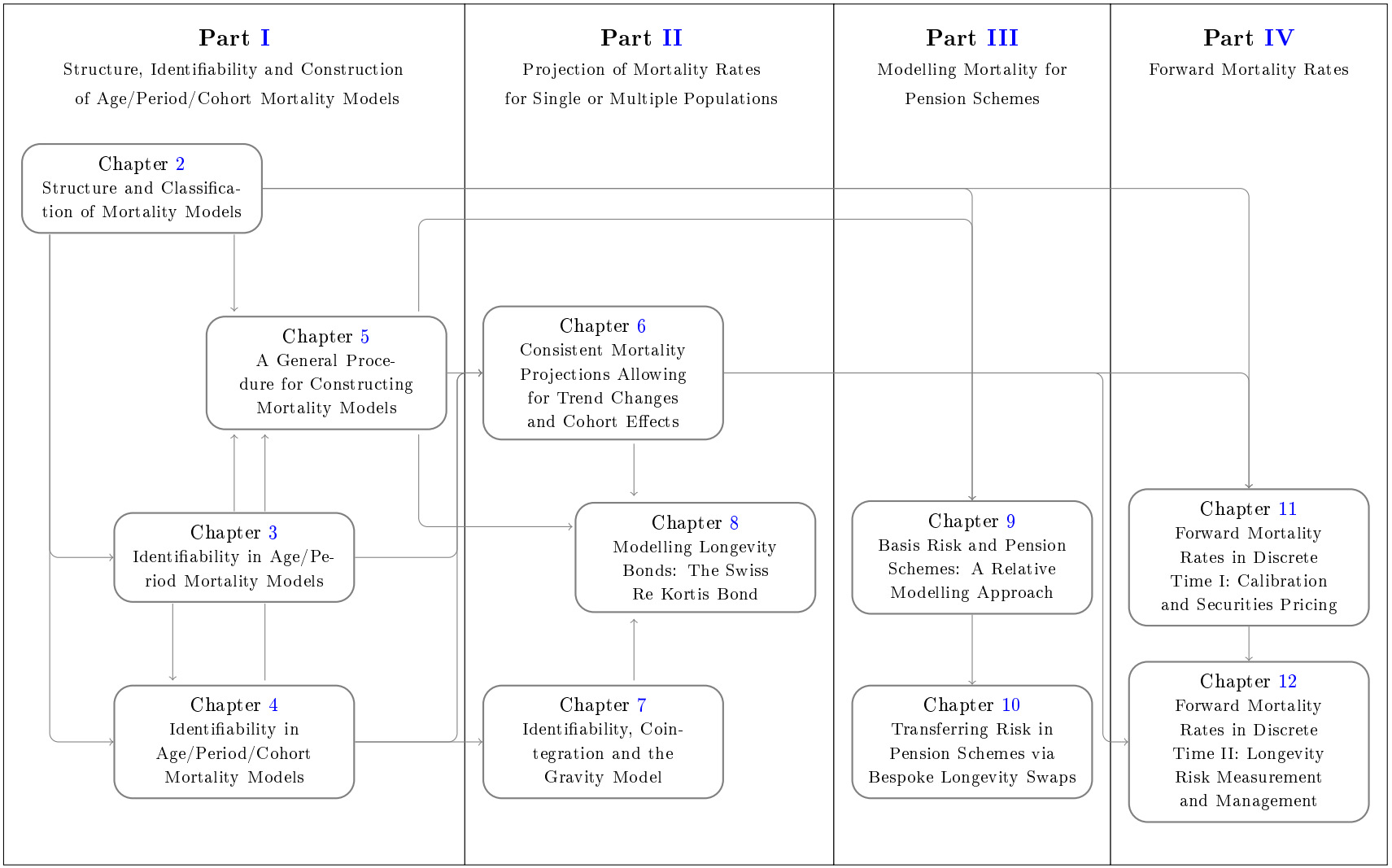


FIGURE 1.1: Dependence structure of chapters in thesis

1.1.2 Identifiability in Age/Period Mortality Models

I am grateful to Andrés Villegas, Steven Haberman, Pietro Millossovich, Bent Nielsen and Ana Debón for their detailed comments regarding this chapter.

As the field of modelling mortality has grown in recent years, the models used to analyse and project mortality rates have grown considerably more sophisticated. However, the number and importance of identifiability issues within mortality models has also grown in parallel with this increased sophistication. This has led both to robustness problems and to difficulties in making projections of future mortality rates. This chapter, therefore, presents a holistic and comprehensive analysis of the identifiability issues in age/period mortality models (i.e., a subset of the class of models discussed in Chapter 2) in order to both understand them better and to finally resolve them. In this chapter, we discuss how these identification issues arise, how to choose identification schemes which aid our demographic interpretation of the models and how to project the models so that our forecasts of the future do not depend upon the arbitrary choices used to identify the historical parameters estimated from historical data. In tandem with Chapter 4, this chapter resolves many of the theoretical and practical issues which have hindered the development of more complicated APC mortality models, and thus is fundamental to the general procedure developed in Chapter 5.

1.1.3 Identifiability in Age/Period/Cohort Mortality Models

I am grateful to Andrés Villegas, Matthias Börger and Bent Nielsen for their detailed comments regarding this chapter.

The addition of a set of cohort parameters to a mortality model can generate complex identifiability issues caused by the collinearity between the dimensions of age, period and cohort, beyond those discussed in Chapter 3. As many modern sophisticated mortality models incorporate cohort parameters, this chapter presents a comprehensive analysis of these identifiability issues and how they can be resolved. To achieve this, we discuss the origin of identifiability issues in general APC mortality models before applying these insights to simple but commonly used mortality models. We then discuss how to project mortality models so that our forecasts of the future are independent of any arbitrary choices we make when fitting a model to data in order to identify the historical parameters. Since the majority of models constructed via the general procedure of Chapter 5

and used in the remainder of this thesis include cohort parameters, the analysis of the identifiability issues in APC mortality models is fundamental to much of the following work, especially the studies in Chapters 6, 7 and 8.

1.1.4 A General Procedure for Constructing Mortality Models

This chapter has been published in the North American Actuarial Journal in 2014, volume 18, issue 1, pages 116-138.

Material in this chapter was presented at the Eighth International Longevity Conference in Waterloo, Canada in September 2012 and the Perspectives on Actuarial Risks in Talks of Young Researchers winter school in Ascona, Switzerland, in January 2013. I am grateful to participants at those conferences and the anonymous referee for their comments.

Many of the more complicated APC mortality models proposed recently suffer from being over-parametrised or are extensions of simpler models where terms have been added in an ad hoc manner which cannot be justified in terms of demographic significance. In addition, poor specification of a model can lead to period effects in the data being wrongly attributed to cohort effects, which results in the model making implausible projections. In this chapter, we present a general procedure for constructing mortality models with the class of APC models discussed in Chapter 2, using a combination of a toolkit of functions and expert judgement. By following the general procedure, it is possible to identify sequentially every significant demographic feature in the data and give it a parametric structural form. We demonstrate using UK mortality data that the general procedure produces a relatively parsimonious model that nevertheless has a good fit to the data. The studies of Chapters 3 and 4 ensure that these models are fully identified and do not suffer from robustness issues when fitted to data. The general procedure is subsequently used to construct the models used in all of the following studies, Chapters 6, 8, 9, 10, 11 and 12, and so is fundamental to most of the following work.

1.2 Projection of Mortality Rates for Single or Multiple Populations

For the majority of practical purposes, we not only need to fit a mortality model to historical data but also to use it to project mortality rates into the future. When doing so, it is important that these projected mortality rates are consistent with the mortality rates observed in the historical data. Furthermore, it is essential that the projected mortality rates are independent of the arbitrary identifiability constraints which were imposed when fitting the model to data. This is an especially large problem in multi-population mortality models and may conflict with a desire for "coherence" between populations, namely that mortality rates in related populations do not diverge. Part II of this thesis proposes new techniques for projecting mortality rates in a single population consistently with historical observations and discusses the issue of identifiability in multi-population mortality models. We then apply these results to the modelling of the first "longevity trend bond": the Kortis bond issues by Swiss Re in 2010.

1.2.1 Consistent Mortality Projections Allowing for Trend Changes and Cohort Effects

Material in this chapter was presented at the 17^{th} *International Congress on Insurance: Mathematics and Economics in July 2013 in Copenhagen, Denmark. I am grateful to participants at that conference and to Matthias Börger, Frank van Berkum, Michele Bergamelli and Andrés Villegas for their comments and to Robert Cowell for discussions regarding the Bayesian approach to modelling cohort parameters.*

The extrapolative approach to projecting mortality has the core assumption that there is consistency between the evolution of mortality rates in the past and the future. When using extrapolative mortality models, there is therefore a fundamental symmetry between the processes of fitting the model to historical observations to find parameter estimates, on the one hand, and projecting parameter values to project future observations, on the other. Consequently, it is important that the models we use to project mortality genuinely achieve consistency between the past and the future. This chapter proposes a number of new techniques to project mortality consistently using the APC mortality model developed in Chapter 5, both across periods, by allowing for observed and future trend changes, and along cohorts, by allowing for the limited observations we have to date for those cohorts that are still alive. Care is taken to ensure that these projections are independent of the arbitrary identifiability constraints imposed in order to resolve

the identifiability issues present in the models, discussed in Chapters 3 and 4. When using these techniques, we obtain projections which are closer to observed mortality rates when backtested and are more biologically reasonable in the long term compared with standard techniques. In addition, the approach used to model and project the cohort parameters is used in Chapter 8 and extended as part of the forward mortality framework in Chapters 11 and 12.

1.2.2 Identifiability, Cointegration and the Gravity Model

I am grateful to Bent Nielsen and Michele Bergamelli for discussions regarding identifiability and cointegration, which informs the material in this chapter.

For many purposes, it is necessary to be able to project mortality rates in related populations, maintaining any correlations observed in the historical data in our projections of the future. As an example of this, the gravity model of Dowd et al. (2011b) was introduced in order to achieve coherent projections of mortality between two related populations. However, this model as originally formulated is not well-identified, since it gives projections which depend on the arbitrary identifiability constraints imposed on the underlying mortality model when fitting it to data. In this chapter, we discuss how the gravity model can be modified to give well-identified projections of mortality rates and how this result can be generalised to more complicated mortality models, such as those used in Chapter 8.

1.2.3 Modelling Longevity Bonds: The Swiss Re Kortis Bond

This chapter has been published in Insurance: Mathematics and Economics in 2015, volume 63, pages 12-39.

Material in this chapter was presented at the Ninth International Longevity Conference in Beijing, China, in September 2013 and at an internal seminar at Cass Business School in April 2014. I am grateful to participants at those events, to Bent Nielsen and Michele Bergamelli for discussions regarding identifiability and cointegration, and to Daniel Harrison FIA at Swiss Re and the anonymous referee for their detailed comments .

A key contribution to the development of the traded market for longevity risk was the issuance of the Kortis bond, the first longevity trend bond, by Swiss Re in 2010. We analyse the design of the Kortis bond, develop suitable mortality models using the general procedure of Chapter 5 and the projection techniques developed in Chapters 6 and 7 to analyse its payoff and discuss the key risk factors for the bond. We also investigate how the design of the Kortis bond can be adapted and extended to further develop the market for longevity risk.

1.3 Modelling Mortality for Pension Schemes

Much of the research to date has been motivated by the impact of longevity risk on the providers of retirement benefits, which has become especially apparent for occupational pension schemes in the UK. However, many of the sophisticated mortality models developed in the previous parts of this thesis are not appropriate for use with a pension scheme, since they generally possess far more limited data. Furthermore, it is not clear that mortality rates in small sub-populations, such as a pension scheme, will evolve in the same manner as those in a larger reference population. Part III of this thesis, therefore, develops a "relative" mortality modelling approach, which can combine the advantages of using sophisticated mortality models for a reference population, with the need for parsimony and robustness when investigating how mortality rates in a sub-population differ from this reference population. This is then applied to investigate the potential effectiveness of a bespoke longevity swap in hedging the mortality and longevity risks in a stylised pension scheme, typical of those found in the UK.

1.3.1 Basis Risk and Pension Schemes: A Relative Modelling Approach

I am grateful to Andrés Villegas for many useful discussions around the topic of relative modelling and basis risk, which informs the material in this chapter.

For many pension schemes, a shortage of data limits the ability to use sophisticated stochastic mortality models such as those constructed by the general procedure of Chapter 5 to assess and manage their longevity risk. In this chapter, we develop a relative model for mortality, which compares the evolution of mortality rates in a sub-population with that observed in a larger reference population. We apply this relative approach to data from the CMI Self-Administered Pension Scheme study, using UK population data

as a reference, for which we can use more sophisticated models. We then use the relative approach to investigate the potential basis risk between these two populations and find that, in many practical situations, much of the concern regarding basis risk is misplaced. These results are then developed further in Chapter 10.

1.3.2 Transferring Risk in Pension Schemes via Bespoke Longevity Swaps

The pensions de-risking industry has grown enormously in recent years, with the focus of much of this on transferring the mortality and longevity risks of pension schemes to third parties. Bespoke longevity swaps, tailored to the specific characteristics of the transferring scheme, have been developed to transfer these risks to insurers and reinsurers and have proved very popular, with over £50bn of outstanding deals transacted to Q4 2014. In this study, we present a modelling framework suitable for assessing the various mortality and longevity risks within a stylised pension scheme and use this to give a comprehensive analysis of the mortality and longevity risks in a pension scheme, and hence the effectiveness of a bespoke longevity swap. In particular, we focus on the possible interactions between the different risk factors that influence mortality rates. This uses a model developed for the national population using the general procedure of Chapter 5 to incorporate systematic longevity risk, the relative model developed in Chapter 9 to model basis risk between the national population and the scheme and also allows for individual mortality effects to give a more complete analysis of the mortality and longevity risks in a scheme, and hence the effectiveness of a bespoke longevity swap in reducing the risk faced by a pension scheme.

1.4 Forward Mortality Models

When valuing longevity-linked liabilities and securities, we are interested in what our expectations of future mortality rates are, conditional on the information we have to date. Where market prices are available, these inform our expectations and ensure that our values are consistent with the values of traded securities existing in the market. To do so efficiently requires the use of a forward mortality model. Furthermore, to measure longevity risk, we need a forward mortality model capable of assessing how the values of longevity-linked liabilities and securities change in response to new information. In Part IV of my thesis, we develop a new forward mortality framework, which builds

on the structure of APC mortality models. This framework is capable of valuing and measuring the longevity risk present in longevity-linked liabilities and securities, which has applications for the new Solvency II regulatory standards and the hedging of longevity risk using simple longevity-linked securities.

1.4.1 Forward Mortality Rates in Discrete Time I: Calibration and Securities Pricing

Material in this chapter was presented at the 49th Actuarial Research Conference in Santa Barbara, USA, in July 2014, the Tenth International Longevity Conference in Santiago, Chile in September 2014, and the Society of Actuaries Longevity Seminar in Chicago, USA, in February 2015. I am grateful to participants at those conferences for their comments.

Many users of mortality models are interested in using them to place values on longevity-linked liabilities and securities. Modern regulatory regimes require that the values of liabilities and reserves are consistent with market prices (if available), whilst the gradual emergence of a traded market in longevity risk needs methods for pricing new types of longevity-linked securities quickly and efficiently. In this chapter and Chapter 12, we develop a new forward mortality framework to enable the efficient pricing of longevity-linked liabilities and securities in a market-consistent fashion. This approach starts from the historical data on the observed mortality rates, i.e., the observed force of mortality. Building on the dynamics of models of the observed force of mortality, we develop models of forward mortality rates and then use a change of measure to incorporate whatever market information is available. This framework is applicable for most models within the class of APC mortality models discussed in Chapter 2, including those constructed using the general procedure of Chapter 5 and uses the Bayesian approach to model and project the cohort parameters developed in Chapter 6.

1.4.2 Forward Mortality Rates in Discrete Time II: Longevity Risk Measurement and Management

Material in this chapter was presented at the 49th Actuarial Research Conference in Santa Barbara, USA, in July 2014, the Tenth International Longevity Conference in Santiago, Chile in September 2014, and the Society of Actuaries Longevity Seminar in Chicago, USA, in February 2015. I am grateful to participants at those conferences for their comments and to Robert Cowell for discussions regarding the Bayesian approach to modelling

cohort parameters.

It is vital to be able to measure and manage the risk in longevity-linked liabilities and securities reliably and consistently, especially in the context of the rapidly expanding market for longevity risk transfer. In this chapter, we develop the forward mortality framework of Chapter 11 and use it for the measurement of longevity risk in portfolios of annuities and in various longevity-linked securities. This involves extending the method used to model and project the cohort parameters developed in Chapter 6 in order to allow for the impact of new information on our re-estimation of the parameters. We then apply the framework to the hedging of longevity risk using simple longevity-linked securities and as an internal model for longevity risk for the calculation of the Solvency Capital Requirement and the Risk Margin under the forthcoming Solvency II regulatory standards.

Part I

Structure, Identifiability and Construction of Age/Period/Cohort Mortality Models

Chapter 2

On the Structure and Classification of Mortality Models

2.1 Introduction

Recent years have witnessed a dramatic increase in the attention paid to the study of the evolution and projection of mortality rates. Demographers, statisticians and actuaries across the world have woken up to the issues caused by rising longevity and an aging population.

Much of the analysis of the historical evolution of mortality rates is made using models which decompose mortality rates across the dimensions of age, period and cohort (or year of birth). These three variables form a natural way of analysing how mortality rates change for individuals as they age, the impact of medical and social progress with time, and the lifelong mortality effects which follow individuals from birth. By projecting the effects of period and cohort, we can also gain insights into the likely path mortality rates might take in future.

Since the number of age/period/cohort (APC) models has increased rapidly in recent years, we believe that the time has come to undertake a more holistic analysis of APC models. We do this in a series of studies, of which this is the first. This present chapter analyses the structure of APC models and proposes a way of classifying the models proposed to date. It also seeks to assess the key principles a model user should consider before selecting or constructing a model appropriate to their aims. While most of the issues raised in this study will be familiar to many model users, we believe that a proper

understanding of the structure of APC models is needed in order to avoid using a poorly specified model. As well as using models which are not suitable for the task in hand, a poorly chosen APC model might also suffer from problems both with the identifiability of parameters in-sample and with projections out of sample. These issues are dealt with in our second and third studies, Chapters 3 and 4. Many of the issues raised and pitfalls identified in these studies were vital to the development of the "general procedure" for constructing APC mortality models, described in Chapter 5.

We discuss the basic structure of the majority of APC models which have been proposed to date in Section 2.2. The components of this structure are further discussed in terms of

- the connections between the data, the variables of interest and our predictor structure in Section 2.3;
- the inclusion of a static function of age in Section 2.4;
- the potential forms for the dynamic structure across ages in the model in Section 2.5; and
- the issues raised by the inclusion of parameters to capture the effects of year of birth in the data and how these can be resolved in Section 2.6.

Section 2.7 offers a simple classification of APC models that highlights the key decisions which have to be made in order to select the most suitable model for the task at hand. Finally, we draw conclusions in Section 2.8.

2.2 Age/period/cohort structure

An APC mortality model is one which links a response variable with a linear or bilinear predictor structure consisting of a series of factors dependent on age, x, period, t, and year of birth (or cohort), $y = t - x$, for a population. APC models therefore fit into the general class of generalised non-linear models, with a general structure which can be written as follows:

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_{t-x} \tag{2.1}$$

This structure has the following components:

- A link function, $\eta_{x,t}$, to transform the response variable (which will be some measure of mortality rates) at age x and for year t into a form suitable for modelling and link it to the proposed predictor structure.
- A static age function, α_x, to capture the general shape of mortality across all ages and features of the mortality curve which do not change with time.
- A set of N age/period terms, $\beta_x^{(i)}\kappa_t^{(i)}$, consisting of period functions, $\kappa_t^{(i)}$, determining the evolution of mortality rates through time, and age functions, $\beta_x^{(i)}$, determining the pattern of mortality change across ages. The choice of suitable forms for the age functions is discussed in Section 2.5.
- An age/cohort term, $\beta_x^{(0)}\gamma_{t-x}$, consisting of a cohort term, γ_{t-x}, which determines the lifelong effects specific to each generation, denoted by their year of birth, and an age function, $\beta_x^{(0)}$, which modifies the cohort term.[1]

Each of these component terms is discussed in greater detail in the sections below. One advantage of most APC mortality models is that the components in them can be interpreted in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them. We call such an interpretation the "demographic significance" of each term. Demographic significance is, by definition, subjective as it relates to the interpretation of the parameters. However, it is still a useful concept as it motivates many of the decisions around the construction of mortality models and their projection into the future.

While this structure is not exhaustive, it does encompass the vast majority of the discrete time mortality models which have been proposed to date. In particular, it is worth noting that we have assumed that the period functions can vary freely for each year and are not constrained to be smooth functions. This is the key feature which enables these models to be projected stochastically and therefore generate probabilistic forecasts of future mortality rates.

In contrast, some models, such as the P-splines model proposed by Currie et al. (2004) and the model of Sithole et al. (2000), require that the period functions be modelled through a series of basis functions (cubic b-splines and Legendre polynomials, respectively) and so are projected by extrapolating these deterministic functions into the future. This typically restricts the application of these models to smoothing historical data or

[1] Most APC mortality models have only one age/cohort term for the reasons discussed in Section 2.6. However, some models do incorporate multiple terms, for instance, that proposed in Hatzopoulos and Haberman (2011).

short-term projections of mortality. We therefore do not consider these models further in this chapter.

Recently, a number of studies, such as Mitchell et al. (2013), Haberman and Renshaw (2012) and Haberman and Renshaw (2013), have modified the structure in Equation 2.1 to model mortality improvement rates rather than the mortality rates themselves. The different interpretations placed on the response variables of interest and terms within the predictor structure make mortality improvement models qualitatively different from the class of models considered within this study, and so we do not discuss these models further.

Finally, it is worth noting that the predictor structure in Equation 2.1 could also be extended to include a range of explanatory variables which might influence mortality rates. These regressors might include variables relating to the health of the population (for instance, smoking prevalence was considered in Wang and Preston (2009) and Kleinow and Cairns (2013)) or macroeconomic variables such as GDP growth or unemployment (e.g., Reichmuth and Sarferaz (2008) and Hanewald (2011)). Such an approach is a natural way of modelling the underlying drivers of changing mortality and highlights the flexibility of the APC approach, but is again not considered further in this chapter.

2.3 Response variable and link function

When studying mortality, we typically assume that members of the population of interest experience the same instantaneous hazard rate of mortality, $\mu_{x,t}$, at age x and time t (also called the "force of mortality"). In practice, however, observed data is usually grouped into discrete age and period bands and therefore modelling mortality is often conducted using discrete time models.

In order to use the continuous force of mortality in a discrete age/period setting, it is commonly assumed that mortality rates do not change within each age and period band. Mathematically, this means that $\mu_{x,t}$ is assumed to be constant within ages and within years:

$$\begin{aligned} \mu_{x+\xi,t+\tau} &= \mu_{x,t} \\ x, t &\in \mathbb{N} \\ \xi, \tau &\in [0,1) \end{aligned} \tag{2.2}$$

This assumption is generally reasonable for most ages of interest (typically under age 100). Above this age, the populations under observation and correspondingly the number of deaths tend to be quite low, which means that the practical impact of this assumption breaking down is quite small over most ages. With the assumption that the force of mortality is constant over each age/period band, we therefore have that the probability of survival over the period is $p_{x,t} = 1 - q_{x,t} = \exp(-\mu_{x,t})$ and that the central mortality rate is given by $m_{x,t} = \mu_{x,t}$. Almost all APC mortality models either use $\mu_{x,t}$ (or equivalently $m_{x,t}$) or $q_{x,t}$ as the response variable for mortality.

These two choices for the response variable reflect the two models for the random number of deaths, $D_{x,t}$, widely used in demography and actuarial science. Under the binomial assumption, the expected number of deaths is given by $\mathbb{E}(D_{x,t}) = E^0_{x,t} q_{x,t}$, the initial number of people alive (or initial exposure to risk) multiplied by the probability of death over the year. The probability of death can therefore be estimated as the observed number of deaths divided by the initial exposure to risk, $\hat{q}_{x,t} = \frac{d_{x,t}}{E^0_{x,t}}$.[2] Under the Poisson assumption, the expected number of deaths is given by $\mathbb{E}(D_{x,t}) = E^c_{x,t} m_{x,t}$, i.e., the central exposure to risk (the average number of people alive which is used as a proxy for the total number of person-years lived) multiplied by the central mortality rate, $\hat{m}_{x,t} = \frac{d_{x,t}}{E^c_{x,t}}$.

This leads to the conclusion that the model for the response variable should be motivated by the format of the available data. The use of the Poisson model requires central exposures to risk which are widely available, for instance from the Human Mortality Database.[3] The use of the binomial model requires initial exposures to risk which are less commonly available for large populations (though may be more available for smaller populations) but can be approximated from the central exposures.

Asymptotically, for large populations and low death rates, the two approaches give similar results. It has been argued[4] that the binomial approach works well at high ages, since it gives transformed mortality rates which are closer to being linear at the highest ages. However, it is also at these ages that the assumption of a constant force of mortality within ages and years in Equation 2.2 starts to break down. Since this violates the core assumption underpinning the discrete time approach, it means that the validity of all models become questionable at these ages and hence makes comparisons between them

[2] Where $d_{x,t}$ is the observation of the random death count, $D_{x,t}$.
[3] Human Mortality Database (2014).
[4] For instance, in Cairns et al. (2006a).

at these ages somewhat spurious.[5]

In the Poisson and binomial models, the variances of the observations are also specified along with the means. In practice, however, observations typically show a greater variation than is predicted under either distribution - a phenomenon known as over-dispersion. One way of dealing with this is by fitting the model using the quasi-Poisson or quasi-binomial distributions, which add additional parameters to account for the over-dispersion. Alternatively, heterogeneity and over-dispersion within the data can be allowed for by using the negative binomial model for death counts, as in Delwarde et al. (2007b), Renshaw and Haberman (2008) and Li et al. (2009). These approaches do not change the model structure in Equation 2.1, merely how it is fit to data. However, over-dispersion (along with significant correlation patterns within the fitted residuals) may also be a sign that the predictor structure is poorly chosen and so could be dealt with by selecting an alternative predictor structure.

The link function, $\eta_{x,t}$, provides the connection between the observed data and the assume predictor structure. In the generalised linear model framework, there are several requirements which should be met for a good choice of link function. One of these is that the data should be transformed to obtain an approximately linear predictor structure (as opposed to, say, a multiplicative structure). Early static and dynamic mortality models used this as the sole requirement for the choice of $\eta_{x,t}$, which resulted in a range of choices being made, such as $\eta_x = \frac{q_x}{1-q_x}$ in Heligman and Pollard (1980), $\eta_{x,t} = \ln\left(\frac{q_{x,t}}{1-0.5q_{x,t}}\right)$ in Wilmoth (1990) and $\eta_{x,t} = \ln(\mu_{x,t})$ in Lee and Carter (1992). These models were then fitted using least squares estimation methods.

Least squares methods, however, do not account for the underlying distribution for $D_{x,t}$ and assume that the variance of observations is independent of the underlying exposures. However, this is not usually valid - observations are typically more variable at ages with low populations, such as those at high ages. More sophisticated methods of estimation, based on maximising the likelihood (Brouhns et al. (2002a)) or, equivalently, minimising the scaled deviance (Renshaw and Haberman (2003a)) allow for this directly by making explicit reference to the underlying probability distribution of $D_{x,t}$. Although a number of potential link functions might be considered for either distribution of death counts (for instance, see Currie (2014)), practical considerations motivate using the canonical link function of the distribution $D_{x,t}$. The choice of the canonical link function also ensures

[5] One solution to this might be to assume a constant force of mortality over shorter age and period bands, for instance across months as in Gavrilov and Gavrilova (2011). However, data limitations at high ages often prevent this.

that fitted values of the response variable lie within the required range.[6] For a Poisson model of the death count, the canonical choice for the link function $\eta_{x,t}$ is

$$\begin{aligned} \eta_{x,t} &= \ln(\mu_{x,t}) \\ \mathbb{E}[D_{x,t}] &= E^c_{x,t} e^{\eta_{x,t}} \\ \mathbb{V}ar(D_{x,t}) &= E^c_{x,t} e^{\eta_{x,t}} \end{aligned} \tag{2.3}$$

whilst for the binomial model it is

$$\begin{aligned} \eta_{x,t} &= \text{logit}(q_{x,t}) \equiv \ln(q_{x,t}) - \ln(1 - q_{x,t}) \\ \mathbb{E}[D_{x,t}] &= E^0_{x,t} \frac{e^{\eta_{x,t}}}{1 + e^{\eta_{x,t}}} \\ \mathbb{V}ar(D_{x,t}) &= E^0_{x,t} \frac{e^{\eta_{x,t}}}{(1 + e^{\eta_{x,t}})^2} \end{aligned} \tag{2.4}$$

Using the canonical link function also has the desirable property that it simplifies estimation by maximum likelihood on minimal deviance considerably easier. For Poisson death counts using the log link function, the likelihood function is

$$\mathcal{L} = \sum_{x,t} W_{x,t} \left(d_{x,t} \ln(E^c_{x,t} \mu_{x,t}) - E^c_{x,t} \mu_{x,t} - \ln(d_{x,t}!) \right) \tag{2.5}$$

whilst for binomial death counts and the logit link function, the likelihood function is

$$\begin{aligned} \mathcal{L} = \sum_{x,t} W_{x,t} \Big(& d_{x,t} \ln(q_{x,t}) + (E^0_{x,t} - d_{x,t}) \ln(1 - q_{x,t}) \\ & + \ln(E^0_{x,t}!) - \ln((E^0_{x,t} - d_{x,t})!) - \ln(d_{x,t}!) \Big) \end{aligned} \tag{2.6}$$

where $W_{x,t}$ are $\{0, 1\}$ weights. When using Newton-Raphson techniques to maximise the likelihood, we need to calculate the first and second derivatives of the log-likelihood function with respect to the parameters (e.g., see Brouhns et al. (2002a)), the forms of

[6] i.e., $\mu_{x,t} \geq 0$ or $q_{x,t} \in (0, 1)$.

which are

$$\begin{aligned}
\frac{d\mathcal{L}}{d\alpha_x} &= \sum_t \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right) \\
\frac{d^2\mathcal{L}}{d\left(\alpha_x\right)^2} &= -\sum_t \mathbb{V}ar(D_{x,t}) \\
\frac{d\mathcal{L}}{d\beta_x^{(i)}} &= \sum_t \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right)\kappa_t^{(i)} \\
\frac{d^2\mathcal{L}}{d\left(\beta_x^{(i)}\right)^2} &= -\sum_t \mathbb{V}ar(D_{x,t})\left(\kappa_t^{(i)}\right)^2 \\
\frac{d\mathcal{L}}{d\kappa_t^{(i)}} &= \sum_x \left(d_{x,t} - \mathbb{E}[D_{x,t}]\right)\beta_x^{(i)} \\
\frac{d^2\mathcal{L}}{d\left(\kappa_t^{(i)}\right)^2} &= -\sum_x \mathbb{V}ar(D_{x,t})\left(\beta_x^{(i)}\right)^2 \\
\frac{d\mathcal{L}}{d\gamma_y} &= \sum_x \left(d_{x,x+y} - \mathbb{E}[D_{x,x+y}]\right)\beta_x^{(0)} \\
\frac{d^2\mathcal{L}}{d(\gamma_y)^2} &= -\sum_x \mathbb{V}ar(D_{x,x+y})\left(\beta_x^{(0)}\right)^2
\end{aligned}$$

These are simple to compute quickly if the canonical link is used. Alternative link structures require more complicated algorithms[7] which it may be desirable to avoid.

Any decisions regarding the choice of response variable and link function should take the following into account:

- The choice of probability distribution should reflect the available data - the binomial distribution is the natural choice with initial exposures to risk, whilst the Poisson distribution is more natural for model users with central exposures.
- The choice of response variable follows naturally from the probability distribution - $\mu_{x,t}$ is the variable of interest in the Poisson distribution and $q_{x,t}$ in the binomial distribution.
- The appropriate canonical link function $\eta_{x,t}$ follows naturally from the probability distribution selected. While other link functions can be chosen, such a choice would probably require further justification.

[7] See, for instance, the estimation of models in the CBD family using the LifeMetrics code in Coughlan et al. (2007a), where a Poisson distribution of deaths is assumed with a logit link function.

In practice, most modellers use the $\ln(\mu_{x,t})$ approach, i.e., a log link function, and assume the death count is a Poisson random variable. These models include those proposed in Brouhns et al. (2002a), Renshaw and Haberman (2003b, 2006), Plat (2009a), Haberman and Renshaw (2009) and O'Hare and Li (2012a). However, the reasons for this are mainly historical, as they are based on the model of Lee and Carter (1992) where the log link function was chosen simply to obtain a linear predictor structure rather than with reference to the underlying distribution of the death counts or the available data. The alternative $\text{logit}(q_{x,t})$ approach has mainly been adopted by the Cairns-Blake-Dowd (CBD) family of mortality models (Cairns et al. (2006a) and the extensions of this model in Cairns et al. (2009)),[8] and also in Aro and Pennanen (2011)..

2.4 Static age function

A static age function, α_x, has been used in many mortality models from Hobcraft et al. (1982) and Lee and Carter (1992) onwards. By construction, this captures the features of the mortality curve across the age range of the data which do not change with time. A typical example of such a function, from the Lee-Carter (LC) model (see Section 2.5.1) fitted to male data from the USA (downloaded from the Human Mortality Database (2014)) for the period 1933 to 2007, is shown in Figure 2.1. Across the full age range, this shows features such as the excess number of deaths due to infant mortality at very low ages and accidents at young adult ages, which are common across both time periods and countries.

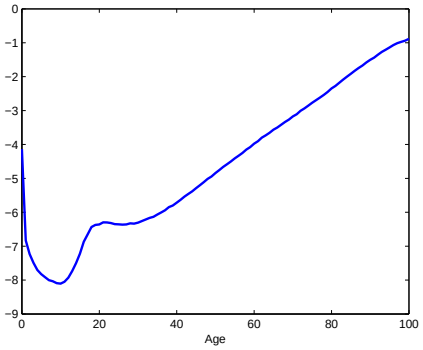


FIGURE 2.1: α_x static age function for the LC model fitted to US male data 1933-2007

Some models, most notably those in the CBD family of mortality models and that in Aro and Pennanen (2011), dispense with the need for an explicit static age function by implicitly assuming that it can be approximated by a simpler function of age and

[8] These models do not draw a direct link between the use of the logit function and binomial death counts. However, this connection is made explicit in Haberman and Renshaw (2011) and Currie (2014).

combining it into the age/period terms. To do this, the static age function needs to be a linear combination of the other age functions in the model, i.e.,

$$\alpha_x = \sum_{i=1}^{N} \alpha^{(i)} \beta_x^{(i)}$$

This can only be done when the age functions $\beta_x^{(i)}$ are known in advance of fitting the model to data. For example, the model of Cairns et al. (2006a) implicitly assumes that mortality rates are approximately linear at the ages of interest and therefore can be combined with the other terms in the model.

Doing so improves the parsimony of the model by reducing the number of free parameters considerably. However, it does so at the expense of limiting the model to only those parts of the age range where this assumption is approximately valid, typically at higher ages.

It also means that the age/period terms in the model do two tasks simultaneously: capturing the time-independent shape of mortality and describing the structure of the deviations from this shape. Including a static age function in the model therefore allows each term in the model to focus on doing one job optimally. The extent to which this is desirable will depend upon the modeller's preference for a parsimonious fit to historical data against the more detailed identification and projection of evolving trends.

2.5 Age/period terms

The age/period terms in an APC model typically capture the majority of the dynamic structure present in the underlying data. They consist of age functions, $\beta_x^{(i)}$, describing how the particular mortality effects are distributed across ages, which are multiplied by period functions, $\kappa_t^{(i)}$, which explain how they evolve with time.

One of the key distinctions between APC models is whether the age effects are modelled using "non-parametric" or "parametric" age functions. Some mortality models have age functions which are "non-parametric" in the sense that values of $\beta_x^{(i)}$ at different ages, x, are fitted without imposing any a priori structure. Age is treated as an unknown factor in the model rather than a regressor with a known structure.[9] Other mortality models

[9] For this reason, we could alternatively refer to non-parametric age functions as "factorial" age functions.

have age functions which are "parametric", since they take a specific functional form that is defined by an algebraic formula.[10]

We should note that our definitions of the terms "non-parametric" and "parametric" differs from other definitions of these terms used in statistics and actuarial science. For the avoidance of doubt, we use the terms to specifically refer to the structure of the age/period terms, and they have no implication for the methods used to fit the model to data. For example, Haberman and Renshaw (2009) and Haberman and Renshaw (2011) used the term "parametric" to refer to the predictor structure for general APC mortality models, and describe any models within this class as "parametric mortality models". Alternatively, "parametric" can refer to the underlying distributional assumptions for the model and the methods used to fit it to data – as such, the assumption of a Poisson distribution of deaths and maximum likelihood estimation would lead to a "parametric mortality model" under this definition. Our usage of these terms is restricted solely to the form of the age effects.

2.5.1 Non-parametric age functions

Most of the early mortality models used non-parametric age functions, e.g., Lee and Carter (1992) (which had a single age function) and Wilmoth (1990) (which had used a parametric age function for the first age/period term but allowed for non-parametric age functions beyond this). Allowing $\beta_x^{(i)}$ to be non-parametric means that it can take any shape in order to maximise the goodness of fit to the data. Such models are necessarily bilinear, as both age and period are unknown factors.

The simplest model to use non-parametric age functions was that proposed by Lee and Carter (1992). It has a single age/period term of the form

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \tag{2.7}$$

More complicated non-parametric approaches emerge naturally from model fitting techniques based principal component analysis (PCA), often based on singular value decomposition (SVD),[11] although they can easily be deployed in a generalised non-linear

[10] For this reason, these age functions could also be called "formulaic".

[11] As used in Lee and Carter (1992), Wilmoth (1990), Booth et al. (2002), Hatzopoulos and Haberman (2009) and Yang et al. (2010) for example.

modelling or maximum likelihood framework.[12] The non-parametric approach also easily extends to an arbitrary number of age/period terms as in Booth et al. (2002), Renshaw and Haberman (2003b) and Hatzopoulos and Haberman (2009). The number of age/period terms in the model is then selected with reference to the data, rather than having been prescribed in advance.

The main advantage of this approach is that the shapes of the age functions are chosen to maximise the fit to the data. This means that each term extracts the maximum amount of information from the data possible. For example, the terms produced by PCA are ranked in order of information extraction - as measured by the percentage of the total variability in the data explained - which makes it possible to select algorithmically an optimal number of terms in the model.

The non-parametric approach is also very flexible. It can be applied quickly and easily across a variety of datasets, as described, for example, in Tuljapurkar et al. (2000) who use the LC model to fit data from a number of developed nations. Similarly, the non-parametric approach can be used across the full age range, whilst parametric age functions are often only suitable for limited age ranges. It also avoids subjective judgements in constructing the model, as terms are fitted automatically to maximise the fit to data. This ability to objectively pick out the most important structure within the data is used as the starting point for the "general procedure" for constructing mortality models outlined in Chapter 5.

However, non-parametric approaches have a number of downsides. Most importantly, the form of the non-parametric age functions generated usually lack demographic significance. For instance, Figure 2.2 shows the β_x age function produced by fitting the LC model to the same data for men in the US used in Section 2.4. It shows that, over the period, improvements in mortality rates have been far faster at young ages (below 20, but especially at age one) than at higher ages, where improvements have been more evenly distributed across ages. It is very difficult to think of an explanation for this shape which does not involve several drivers of changing mortality rates over the period (such as improved hygiene reducing mortality across all ages, childhood vaccination programmes reducing the number of deaths amongst the very young, and improved treatment of

[12] PCA assumes homogenous, normally distributed residuals and, therefore, is inconsistent with the underlying binomial or Poisson distribution for the death count process. However, the estimates obtained for the parameters using PCA can be used as the starting point for methods such as maximum likelihood which use the death count process to allow for heterogeneity caused by differences in the underlying exposures.

cardio-vascular disease in later life).

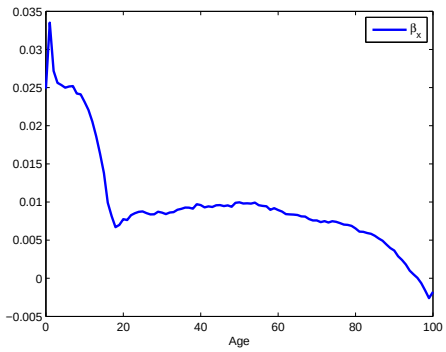


FIGURE 2.2: β_x age function for the LC model fitted to US male data 1933-2007

This has ramifications when we fit and project the model. Drivers of mortality are combined into a single term if they are correlated over the historical period of the data (e.g., they go from a high level of mortality to a lower level over the period). However, these combinations may not be appropriate over subsets of the period range. For example, Carter and Prskawetz (2001) found that the form of β_x changes substantially if the LC model is fitted to different subintervals of the data, as different medical and socio-economic causes of mortality become more or less important.

These combinations of drivers may also be inappropriate when we come to making forecasts using the model. For instance, we may believe that the shape of β_x in Figure 2.2 is due to a combination of childhood immunisation programmes and improved cardio-vascular care for the elderly. When projecting mortality, we may wish to allow the latter to continue to improve in future but believe that we are unlikely to see further reductions in mortality due to increased vaccination of children. Using a term which combines both these causes can lead to projections of mortality rates which do not appear to be plausible, e.g., when high rates of improvement in mortality are projected at ages where mortality rates are already very low.

In addition, the model does not require that the non-parametric forms are continuous.[13] This can lead to projections which have discontinuous mortality rates and so are not biologically reasonable[14] if projected far into the future. It is possible to smooth the

[13] This can be seen with the sharp peak at β_1 in Figure 2.2.

[14] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*". Note that biological reasonableness is a property of observable quantities such as life expectancies or mortality rates, in contrast to demographic significance which relates to our interpretation of the terms in a model.

non-parametric age functions to prevent this, as discussed in Delwarde et al. (2007a) or Hyndman and Ullah (2007). However, this complicates the structure of the model and introduces subjective decisions regarding the degree of smoothing which would need careful justification.

2.5.2 Parametric age functions

As discussed earlier, a parametric age function takes a specific functional form, i.e., $\beta_x = f(x)$. The original APC model, given in Equation 2.8 and first used in the fields of demography, sociology and medical statistics (for instance see Hobcraft et al. (1982)), uses the parametric age functions $\beta_x = f(x) = 1$:

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \tag{2.8}$$

More recently, the CBD model of Cairns et al. (2006a) shown in Equation 2.9 adopts an explicit parametric form (including for the static age function) for both its period functions, with $\beta_x^{(1)} = f^{(1)}(x) = 1$ and $\beta_x^{(2)} = f^{(2)}(x) = (x - \bar{x})$:

$$\text{logit}(q_{x,t}) = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} \tag{2.9}$$

Since the publication of the CBD model, models with increasingly complex parametric age functions have been proposed, such as the extensions to the model in Equation 2.9 in Cairns et al. (2009) and the models proposed in Plat (2009a), Aro and Pennanen (2011), O'Hare and Li (2012a) and Börger et al. (2013).

We can see that the models in Equations 2.8 and 2.9 have a linear predictor structure, rather than possessing any bilinear terms where the age function also needs to be fitted to the data. This means that they are conventional generalised linear models and can be fitted using standard techniques. However, the use of parametric age functions does not necessarily imply linearity. For instance, consider the model

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \exp(-\lambda x)\kappa_t^{(2)}$$

Here, $f^{(2)}(x) = \exp(-\lambda x)$ is parametric in our sense of having a prescribed functional form, but λ can be a free parameter set with reference to the data and so the age/period term is bilinear and the model cannot be estimated via a generalised linear model. Age functions including free parameters are not widely used, as the higher order age functions in the models of Plat (2009a), Aro and Pennanen (2011), O'Hare and Li (2012a) and Börger et al. (2013) have parameters which are set a priori. In principal, however, these

models could be extended to allow these parameters to vary to best fit the data. In addition, many of the age functions used in the "general procedure" of Chapter 5 possess free parameters and therefore are bilinear, parametric age/period terms.

One of the major advantages of using parametric age functions is that they reduce considerably the number of free parameters needing to be fitted for each age/period term, leading to more parsimonious models. This, in turn, means that more parameters can be devoted to detecting other features of interest within the data, such as additional structure across time and year of birth.

Further, because the shapes of the age functions are known, each term can be assigned a specific demographic significance by the user. For instance, the first age/period term in the models of Equations 2.8 and 2.9 are constant across all ages. This can be explained in terms of specific phenomena which are universal across the age range (such as improved hygiene), in contrast with the shape seen in Figure 2.2. It will also allow trends which are correlated (such as improving levels of medical care for the elderly and the specific efforts to tackle childhood infectious diseases) to be given their own age/period terms with appropriate parametric age functions, which is impossible with a non-parametric approach.

However, this flexibility comes at a cost. Parametric age functions are often only suitable over limited age ranges. While this is an advantage in that it allows for greater interpretability of their demographic significance, it means that models with parametric age functions are often not suitable over the full age range. For instance, even if the CBD model were extended with a static age function, it is unlikely that the two age/period terms are sufficient to capture the variability of mortality rates at younger ages. In order to construct a model appropriate across the full age range, we would have to add additional age/period terms to the model.

In addition, models with parametric age functions often give a poorer fit to the data compared to a model with the same number of non-parametric age/period terms, especially using measures of goodness of fit that do not (or only weakly) penalise the number of free parameters in the model. This is because the additional freedom in the non-parametric age function can be used to capture more of the structure in the data than if the form of the age function is prescribed at the outset.

These problems can be rectified, in part, through adding new terms to the model. However, we will need to decide on the appropriate form for these new terms, which can very often be difficult. One approach adopted for some of the extensions to the CBD model in Cairns et al. (2009) is to select age functions from the same family – in this case polynomials of increasing order. Alternatively, more exotic functions can be used as in the models of Plat (2009a) and O'Hare and Li (2012a), but often there does not appear to have any underlying rationale for their selection. In the end, expert judgement is needed to assess whether a new term added to the model genuinely represents the remaining unexplained dominant trend in the data or merely reflects the expectation of the modeller as to what should be present.

2.6 Cohort effects

It is a widely held belief that the different life histories of individuals should lead to systematic difference between people in different cohorts (as summarised by their year of birth). These are often known as "cohort effects". As Hobcraft et al. (1982), Willets (1999) and Murphy (2009) discussed, the term "cohort effect" is largely descriptive, and some care needs to be taken in interpreting the causal factors specific to certain years of birth which might plausibly influence the mortality rate of a cohort across their entire life. We might, for instance, consider an epidemic which, in addition to raising mortality rates at the time it is raging, had a selective effect on the survival of infants. This might lead to systematic differences in mortality between those born during the epidemic and those born shortly before or afterwards. However, the evidence from natural experiments (summarised in Murphy (2009)) is equivocal, which means that the existence of true cohort effects is still controversial to some extent, as discussed in Murphy (2010).

In practice, however, observed data from a number of countries appears to exhibit cohort features and so it is prudent to allow for these when modelling mortality. In the UK, apparent cohort effects have been identified in the general population (specifically in the work of Willets (1999, 2004), Continuous Mortality Investigation (2002) and Richards (2008)) and models allowing for cohort parameters outperformed those which did not in Cairns et al. (2009).

Our subjective demographic significance of a cohort effect is one which increases or reduces mortality at all ages for individuals born in a specific generation (typically lasting 10-15 years or less). To construct a mortality model, we need to translate this demographic significance into a set of properties we desire the parameters in our model to

possess. More specifically, we can say that our intuition regarding the cohort effects implies that they should:

- be small relative to the effects of age and period;
- not have any systematic trends in their expected value or variability;
- have a mean across cohorts of zero (i.e., cohort effects should represent deviations from a typical hypothetical reference level);
- have some autocorrelation: it is reasonable to believe that cohorts born in successive years should experience similar life histories and so exhibit similar cohort effects, unless there happen to be exceptional circumstances facing a particular birth year;
- not exhibit indefinite persistence: the factors influencing the specific mortality of the generation born today should be essentially independent of the specific mortality of their grandparents, for example;
- ideally be mean reverting (as a consequence of the previous two points), as the specific events impacting one cohort wear off in subsequent years of birth; and
- be demographically significant, so we can relate features of a plot of cohort effects to specific socio-economic and medical influences on the population.

In a well-specified mortality model, many of these properties emerge naturally from the fitted parameters. Some, such as the level of the mean of the cohort parameters, can be imposed via identifiability constraints, which change the values of the cohort parameters but not the fit of the model to data. However, this is not always the case, and we may sometimes have to discard some of our intuitive properties based on the evidence of the model. For instance, we can see that in Plat (2009a), the historical cohort parameters have a clear trend and may be non-stationary.

We would also like our cohort parameters to be robust, both across different models and when comparing them with the residuals from the corresponding age/period mortality model, as in Wilmoth (1990). For instance, the plots of cohort parameters for the same datasets in Cairns et al. (2009) show that the features identified are not robust between different models, which weakens any demographic significance we place on them. However, there are a number of practical problems that makes finding cohort parameters that are robust and well specified a harder task than the estimation of age and period parameters.

First, because age, period and cohort are linearly dependent ($y + x = t$), we cannot treat them in isolation for each other.[15] Wilmoth (1990) argued that it is impossible to apportion objectively low frequency (slowly varying) temporal dependence in mortality data between age/period and cohort effects. We therefore are forced to make a subjective choice to give primacy to two of the relevant dimensions. Because we naturally observe cross sections of mortality rates across ages in different calendar years, the data will naturally form a rectangular age/period grid. This means that the natural choice is to give primacy to age and period effects and to try to explain as much of the structure in the data with reference to these dimensions as possible before consideration of effects across cohorts.[16]

This then leads to the conclusion that if the cohort effects are to be taken as of secondary importance, the structure in the model included to capture them should be as simple as possible. Indeed, some have argued that cohort effects do not exist at all and are merely the result of poorly specified age/period effects.[17] A model user operating under such a belief would therefore omit any age/cohort terms from the model entirely. A high standard of evidence for the inclusion of an age/cohort term is therefore desirable.

If an age/cohort term is to be included and if age/cohort interactions are taken to be of secondary importance, the desire for parsimony in the cohort terms leads to two further conclusions which have been adopted by the majority of model users. First, the majority of models only include one cohort term on the grounds that it is hard to believe and to demonstrate that one generation could experience two different independent lifelong effects. Nevertheless, the model proposed in Hatzopoulos and Haberman (2011) allows for multiple cohort effects.

Second, many models set $\beta_x^{(0)} = 1$, leading to a more parsimonious model. This restriction allows the cohort parameters to represent consistently higher or lower mortality rates across all ages, which accords with our demographic interpretation of cohort effects. In particular, while a cohort effect which is stronger at some ages than others does not seem unreasonable in principle, the notion of a cohort effect that increases mortality rates at some ages but decreases them at others conflicts with our interpretation of the demographic significance of a cohort effect. This situation is possible with a non-parametric form for $\beta_x^{(0)}$ unless it is artificially constrained to be greater than zero. In

[15] We also suffer from the problem that the parameters in the model may not be fully identified. This topic and its implications for forecasting are discussed further in Chapter 4.

[16] See Alai and Sherris (2012) for an example of a model which gives primacy to cohort parameters.

[17] For instance, Cairns et al. (2011a) raised *"the possibility that cohort effects might be partially or completely replaced by well-chosen age and period effects"* and also see Murphy (2010)

addition, issues have also been reported concerning the robustness of fitting models such as that of Renshaw and Haberman (2006) with a non-parametric $\beta_x^{(0)}$ term, for instance by Continuous Mortality Investigation (2007) and Cairns et al. (2009).[18] However, this problem is not universal and a linear parametric form for $\beta_x^{(0)}$ was proposed in model M8 (an extension of the CBD model M5 of Equation 2.9) in Cairns et al. (2009) and has been found to be robust and to fit the data well in van Berkum et al. (2014).

Cohort parameters also present specific problems in estimation which again suggests that a parsimonious model structure be used when including them. Because we naturally observe cross sections of mortality rates across ages in different calendar years, we will have a limited numbers of observations for the earliest and latest birth cohorts. This makes estimates of these cohort parameters more uncertain. For instance, the last observed year of birth will only have one observation for it, which can therefore be fit perfectly by the cohort term. This is undesirable and so in practice, many modellers do not estimate cohort parameters for a number of the earliest and latest years of birth in the data (for instance in Renshaw and Haberman (2006) and Cairns et al. (2009)).

Related to this is the fact that the observations for early and late years of birth will only cover a subset of the age range. For instance, the most recent cohorts will only have observations for the youngest ages. Any misspecification of age/period terms affecting these ages will therefore bias the estimation of these cohort parameters. This is especially important for the most recent cohorts, for which we will only have a small number of observations on their early-age mortality where most mortality models have the greatest difficulty modelling the age/period patterns of mortality and where there will be relatively few deaths. Any poorly specified age/period terms at these ages will therefore lead to structure in the data being wrongly attributed to the cohort effect for the most recent years of birth.

As an example of this, there are specific biological factors which lead to mortality in the first year of life evolving differently from mortality rates at subsequent ages. This effect is best captured through an age/period interaction. In a poorly specified age/period mortality model, this cannot be captured adequately, leading to large residuals when fitting mortality rates at this age. Adding a cohort term to such a model will mean that the fitting procedure will try to use the extra parameters to "solve" this problem and so will bias the cohort parameters in order to "fix" what is genuinely an age/period issue. This bias will get more pronounced for more recent years of birth, where observations of

[18] See Hunt and Villegas (2015) for a discussion and potential solution for this issue.

the first year of life form an increasing proportion of the total observations for each new cohort.

In models which give primacy to age/period effects, it is therefore important to ensure that the age/period structure is fully specified before an age/cohort term is added. When forecasting mortality rates, it is of great practical importance that the cohort parameters in an APC model are well specified and estimated robustly. Since cohort effects represent lifelong mortality effects, mis-specifications of the cohort parameters at low ages will bias forecasts for these cohorts as they age.

In summary, the inclusion of a cohort term in a mortality model presents the user with a number of important issues which need to be addressed. In some cases, the model user may consider that cohort effects are not significant and prefer a model which does not include them. However, in other populations, there is evidence to support their inclusion. In such cases, it is necessary to ensure that the age/period structure in the model is well specified and able to capture the majority of structure in the data. A simple and parsimonious cohort term can then be included to capture the effects of year of birth.

2.7 Classification of APC mortality models

Despite the recent rapid proliferation in the number of mortality models proposed, the majority of mortality models in discrete time are part of the same APC family. This then leads to the natural question of how mortality models can be classified.

When constructing an APC mortality model we must ask a number of questions, but especially the following:

- What response variable and link function should we use?
- Should we include an explicit static age function?
- Should we use parametric or non-parametric age functions? If so, how many age/period terms should we use?
- Should we include a cohort term? If so, should it be modified across the age range by a $\beta_x^{(0)}$ age function?

Unlike categorising species of animal, however, our classification of mortality models does not relate to ancestry and so is not unique. What we offer below is a simple classification of mortality models, based on what we consider to be the most important differences in structure between them.

We believe that the first two questions above are straightforward. The modeller's choice for the response variable should depend on the data available to them rather than on any more fundamental consideration. This, in turn, leads to a natural choice for the link function, namely, the canonical link function for the chosen distribution of deaths. Whilst it is possible to use combinations of response variable and link function other than the natural choices, there is often no good reason to do this and practical reasons discussed in Section 2.3 why it should be avoided.

Second, it can be argued that all mortality models use a static age function; it is just that models such as the CBD model of Cairns et al. (2006a) use it implicitly with a distinct parametric structure that enables it to be combined with other terms on the model. Such a choice may be desirable for models limited to specific sections of the age range where the parametric structure is appropriate in order to obtain greater parsimony. However, it does not change anything fundamental about the model.

We are then left with the two more substantive questions - the choice between parametric and non-parametric age functions and the inclusion of a cohort term. Both of these reflect fundamental differences in approach which lead to important mathematical and qualitative differences between the models. Historically, however, cohort parameters have often been seen as an optional addition to a pre-existing mortality model, especially because the age/period terms are usually given primacy due to the reasons discussed in Section 2.6. We, therefore, see the most important division amongst APC models to be between the use of parametric and non-parametric age functions.

The optimum number of age/period terms will then depend on the nature of the age functions chosen to define these terms. In models with non-parametric age functions, it is relatively simple of add additional age/period terms and optimise their number based on a goodness of fit criteria. In models with parametric age functions, however, the number of age functions needs to be defined a priori along with their functional form. If new terms are to be added to an existing model, it is a non-trivial task to select an appropriate form for them. To solve this problem, Chapter 5 introduces a "general procedure" to both select the form of the parametric age functions and determine an optimum

number of age/period terms in a new mortality model.

Based on this analysis, we propose the simple classification of mortality models in Figure 2.3. Obviously this classification is not exhaustive, as new models and variations of existing models are continuously being proposed. It is also not unique, since a different ordering of the questions asked when constructing a mortality model would yield a different family tree. However, we have found it a useful framework when considering the selection of an existing mortality model or when constructing a new one (such as in Chapter 5).

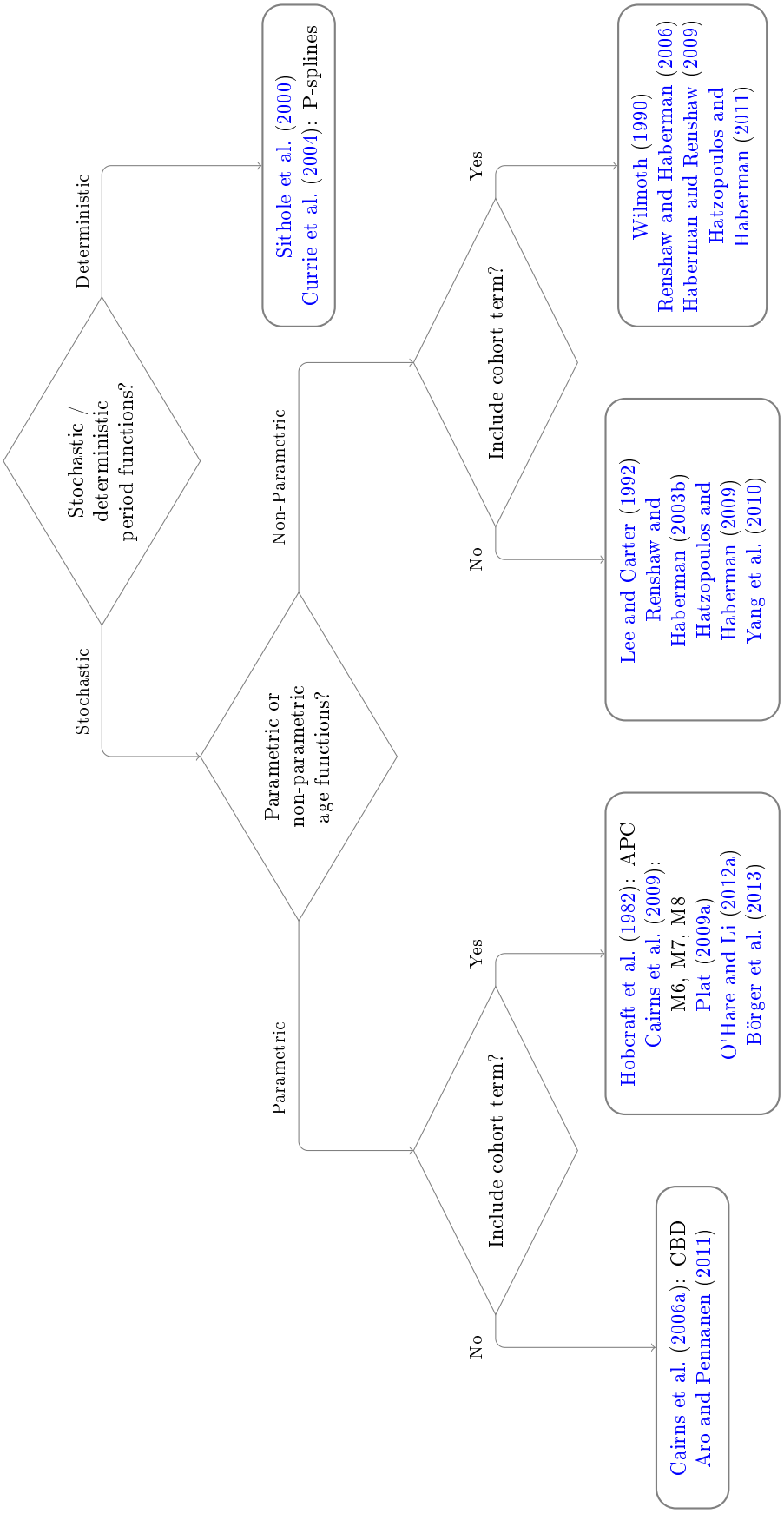


FIGURE 2.3: A simple classification of mortality models

2.8 Conclusions

The increasing number of age/period/cohort models being used to study and project mortality rates has made a general consideration of the APC structure necessary. A systematic and complete understanding of this structure allows us to select or construct the most appropriate model for the dataset and the purpose. We have therefore set out five principles which need to be considered before an APC mortality model can be used or constructed:

1. The response variable being modelled should match the data available. The link functions should follow naturally from the nature of the response variable, e.g., a Poisson distribution for the number of deaths should lead naturally to a log-link function.

2. A static age function should generally be included and made explicit in the model. If a parametric structure is assumed for the static age function, this should be made explicit and the limitations this places on the age range over which the model is suitable should be made clear.

3. The user should justify the choice of a non-parametric or parametric structure for the age functions. Both are appropriate in different circumstances. However, the user of a model should be explicit in the trade-offs they are making between goodness of fit and demographic significance.

4. The use of a cohort term is usually desirable to capture structure across year of birth in the data. However, such a term can be omitted if the evidence does not support its inclusion.

5. When cohort terms are included in a mortality model, they should be made as simple as possible in order to give robust parameter estimates. This will often lead to using a single cohort term and setting $\beta_x^{(0)} = 1$.

We therefore believe that the examination of the structure of APC mortality models in this study has direct practical application when using and developing these models and enables a natural classification to be developed. A proper understanding of the models can therefore help practitioners analyse how mortality has evolved in the past and how it may evolve in future, which is of great importance in the financial and social management of longevity risk in future.

Chapter 3

Identifiability in Age/Period Mortality Models

3.1 Introduction

As the field of modelling mortality has grown in recent years, the models proposed and used have grown ever more complicated. This has had the effect of increasing the number and importance of identifiability issues within the models, which can lead both to robustness problems when fitting the models to data and difficulties when projecting them. As the demands of modern longevity-risk management techniques require sophisticated models capable of capturing complex and subtle relationships between mortality rates across different ages and in different populations, unresolved identifiability issues have important practical consequences. We therefore believe that the time has come for a holistic and comprehensive analysis of the class of age/period/cohort (APC) mortality models and the identifiability issues within them.

In Chapter 2, we analysed the structure of APC mortality models and proposed a way of classifying the models proposed to date. This gave us a general framework in which our study of identifiability issues operates. The existence of identifiability issues means that there are certain features of the parameters in a model which are not defined by the data. Instead, these features are only determined by the arbitrary identifiability constraints we impose upon the model when fitting it to data and, therefore, have no independent meaning. Consequently, we must be careful to ensure that our results from using mortality models do not depend upon these features of the parameters. In the context of the age/period (AP) mortality models discussed in this study, we find that features such as the levels of and correlations between the period terms, and the scale of

the age functions are unidentified by the models. These features therefore do not possess any meaning other than that imposed by our arbitrary identifiability constraints.

Identifiability issues arise in these mortality models because there exist different sets of parameters which will give the same fitted mortality rates. These identifiability issues can lead to models which lack robustness when fitted to data, cause us to draw faulty and erroneous conclusions when analysing the historical data and can bias our projected mortality rates in future. It is essential that we understand and resolve these issues when fitting models to data, as well as comprehend the impact these issues have on our analysis of past and future mortality rates.

Identifiability in mortality models is, therefore, a very important issue. While there are principles which are common to the vast majority of mortality models, the impact and implications of these issues vary considerably depending on the specifics of the model being used. To demonstrate these principles in action, we consider a number of simple models based on the classic and widely used models proposed in Lee and Carter (1992) and Cairns et al. (2006a), both of which are members of the class of AP models. In the particular cases chosen, the identifiability issues can appear trivial, and their impact on our analysis of historical and projected mortality rates relatively minor. However, we believe that it is vital to understand these issues fully in the context of simple models, since they become considerably more important in more sophisticated models, such as those constructed using the "general procedure" of Chapter 5.

In addition, due to the scale of the topic, this study deals only with the identifiability of AP mortality models. We leave the additional issues caused by the inclusion of a cohort term to Chapter 4. Allowing for the dependence of mortality on year of birth in a model often creates new identifiability issues, which are fundamentally different to those affecting simpler AP models and which require a radically different approach to analyse.

We begin, in Section 3.2, by revisiting the general structure of AP models and how identifiability issues arise in them. We then discuss, in Section 3.3, how these issues were dealt with in the model of Lee and Carter (1992), and how this has influenced their treatment in more complex models. The mathematical structure of identifiability issues in the context of these more complex mortality models is investigated in Section 3.4. We then consider how these general issues relate to specific models which are more typical of those used in practice. Section 3.5 discusses the application of the identifiability issues in the context of an extension to the Lee-Carter model. Section 3.6 examines the general

issues in models where the form of the age functions has been chosen a priori. Section 3.7 then considers models which mix age functions of different types.

Identifiability issues in AP mortality models also affect their use in measuring risk and uncertainty in mortality rates. In Section 3.8, we discuss the impact the identifiability issues have on measuring the uncertainty in parameter estimates and on hypothesis testing on the historical parameters. Section 3.9 considers the implications of identifiability issues for projection, and the importance of ensuring that constraints imposed to identify historical parameters uniquely do not impact the projected mortality rates in future. Finally, Section 3.10 concludes.

3.2 Structure and identifiability in age/period mortality models

3.2.1 Structure of age/period mortality models

An AP mortality model in discrete time is one which assumes that mortality rates can be modelled as a series of terms involving functions of age, x, and period, t.[1] In the notation of Chapter 2, this can be written as

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} \tag{3.1}$$

where $\eta_{x,t}$ is a link function transforming the raw data, α_x is a static function of age,[2] $\kappa_t^{(i)}$ are N period functions governing the evolution of mortality with time and $\beta_x^{(i)}$ are age functions modulating the impact of this change over the age range. This structure does not include any allowance for the lifelong effects of different birth years (called "cohort" effects) on mortality.

The structure in Equation 3.1 as it is currently written does not require any of the functions to be known in advance of fitting the model to data. As such, it has what we refer to as a "non-parametric" structure. We consider this as the most general form of an AP mortality model and discuss its identifiability issues in Section 3.4. We will also consider the "parametric" case where $\beta_x^{(i)}$ is a parametric function of age, $\beta_x^{(i)} = f^{(i)}(x; \theta^{(i)})$, in

[1] In this chapter, for generality we assume that $x \in [1, X]$ and $t \in [1, T]$. In practice, the ranges of x and t will be given by the range of the data being used.

[2] Identification issues in models without a static age function, α_x, are discussed in Appendix 3.A.

Section 3.6, and models which mix parametric and non-parametric age functions in Section 3.7. Whether $\beta_x^{(i)}$ is parametric or non-parametric will affect the interpretation of the model, as discussed in Chapter 2, and also lead to subtly different identification issues.

The form given in Equation 3.1 is widely used and lends itself naturally to interpreting the parameters as measuring either an age or a period feature of mortality rates. Alternatively, when analysing this structure, we may find it useful to consider the static age function, α_x, and the age functions, $\beta_x^{(i)}$, as being column vectors in $\mathbb{R}^X$ instead of functions of age, and the period functions, $\kappa_t^{(i)}$, as row vectors in $\mathbb{R}^T$, rather than a time series. Considering the parameters in this context, it is natural to define inner products, $< .,. >$ on $\mathbb{R}^X$ and $\mathbb{R}^T$, respectively, and use these to compare the different functions. For instance, we could define the "scale" of an age function by taking

$$\|\beta_x^{(i)}\| = < \beta_x^{(i)}, \beta_x^{(i)} >$$

or the "angle", θ, between age functions as

$$\cos\theta = \frac{< \beta_x^{(i)}, \beta_x^{(j)} >}{\sqrt{\|\beta_x^{(i)}\|\|\beta_x^{(j)}\|}}$$

We can think of the inner products being the standard Euclidean inner products, i.e. that $< \beta_x^{(i)}, \beta_x^{(j)} > = \sum_x \beta_x^{(i)}\beta_x^{(j)}$ and $< \kappa_t^{(i)}, \kappa_t^{(j)} > = \sum_t \kappa_t^{(i)}\kappa_t^{(j)}$.

If the period parameters, $\kappa_t^{(i)}$, are interpreted as random variables then we also see that this standard Euclidean inner product can be interpreted in terms of their sample mean and sample variance

$$\begin{aligned}
\bar{\kappa}^{(i)} = \frac{1}{T}\sum_t \kappa_t^{(i)} &= \frac{1}{T} < \kappa_t^{(i)}, 1 > \\
\sigma^2_{\kappa^{(i)}} = \frac{1}{T}\sum_t \left(\kappa_t^{(i)} - \bar{\kappa}^{(i)}\right)^2 &= \frac{1}{T} < \kappa_t^{(i)} - \bar{\kappa}^{(i)}, \kappa_t^{(i)} - \bar{\kappa}^{(i)} > \\
&= \frac{1}{T}\|\kappa_t^{(i)} - \bar{\kappa}^{(i)}\|
\end{aligned}$$

Similarly, we see that the sample correlation between two period functions is given by the angle between them

$$\begin{aligned}\text{Corr}(\kappa_{,}^{(i)}\kappa_t^{(j)}) &= \frac{\sum_t \left(\kappa_t^{(i)} - \bar{\kappa}^{(i)}\right)\left(\kappa_t^{(j)} - \bar{\kappa}^{(j)}\right)}{\sqrt{\sigma^2_{\kappa^{(i)}}\sigma^2_{\kappa^{(j)}}}} \\ &= \frac{< \kappa_t^{(i)} - \bar{\kappa}^{(i)}, \kappa_t^{(j)} - \bar{\kappa}^{(j)} >}{\sqrt{\|\kappa_t^{(i)} - \bar{\kappa}^{(i)}\|\|\kappa_t^{(j)} - \bar{\kappa}^{(j)}\|}} \\ &= \cos\theta_{\kappa-\bar{\kappa}}\end{aligned}$$

Consequently, the standard Euclidean inner product has a number of helpful interpretations and is widely used.[3] However, we could equally reasonably choose other inner products on $\mathbb{R}^X$ and $\mathbb{R}^T$ if these are more convenient.[4]

When projecting the period functions using multivariate time series processes, it is helpful to define vectors

$$\begin{aligned}\boldsymbol{\kappa}_t &= \left(\kappa_t^{(1)}, \quad \ldots \quad \kappa_t^{(N)}\right)^\top \\ \boldsymbol{\beta}_x &= \left(\beta_x^{(1)}, \quad \ldots \quad \beta_x^{(N)}\right)^\top\end{aligned}$$

The model therefore has the vector structure

$$\eta_{x,t} = \alpha_x + \boldsymbol{\beta}_x^\top \boldsymbol{\kappa}_t \tag{3.2}$$

In order to project the model, the vector $\boldsymbol{\kappa}_t$ can be modelled using VARIMA processes. This is considered further in Section 3.9.

We can also construct matrices for the age and period functions as $\beta = \{\beta_x^{(1)}\ \beta_x^{(2)} \ldots \beta_x^{(N)}\}$ and $\kappa = \{\kappa_t^{(1)}; \kappa_t^{(2)}; \ldots; \kappa_t^{(N)}\}$ and therefore re-write Equation 3.1 in matrix form

$$H = \alpha 1^\top + \beta\kappa \tag{3.3}$$

where

- H is the $(X \times T)$ matrix of transformed data (i.e., $H = \{\eta_{x,t}\}$),
- α is a $(X \times 1)$ matrix of the static age function,

[3] For example, it is common to impose $\bar{\kappa}_t^{(i)} = \frac{1}{T} < \kappa_t^{(i)}, 1 >= \frac{1}{T}\sum_t \kappa_t^{(i)} = 0$ as an identifiability constraint, as discussed below.

[4] For instance, in Chapter 5, we use the standard $L(2)$ inner product to define orthogonality between age and period functions, but use the $L(1)$ norm to define a normalisation scheme.

- $\mathbf{1}$ is a $(T \times 1)$ matrix of ones, and
- β and κ are the $(X \times N)$ matrix and $(N \times T)$ matrix of age and period functions constructed above, respectively.

When expressed in this form, AP models can be analysed through the prism of matrix algebra and linear mathematics. Specifically, we can see that an AP mortality model is a mapping, Θ, from the space of parameters to the model space, $\mathcal{M}$, of fitted mortality rates.

$$\Theta(\alpha_x, \beta_x^{(i)}, \kappa_t^{(i)}) : \mathbb{R}^X \times \mathbb{R}^{NX} \times \mathbb{R}^{NT} \to \mathcal{M} \subset \mathbb{R}^{X \times T} \tag{3.4}$$

Analysing AP mortality models as linear transformations can be very useful, and is pursued in Sections 3.2.2 and 3.4 and in Appendix 3.B. However, whilst such an abstraction can be useful for some purposes, it is important to remember that the parameters in the model have specific interpretations, for instance, that the period functions are ordered chronologically, and so the problem of identifiability should not be seen purely as an exercise in linear mathematics.

3.2.2 Identifiability in age/period models

An AP mortality model cannot, in general, be estimated as it stands. This is because any parameter estimates would not be unique, since Equation 3.3 is not, in general, fully identifiable.

A model is fully identified when all the parameters in it can be uniquely determined by reference to the available data. In contrast, most mortality models are not fully identified - there exist different sets of parameters which will give the same fitted mortality rates and consequently the same goodness of fit. Although this phenomenon is not unique to mortality models, it is very widespread in mortality modelling and has significant implications when we come to project these models.

The models are not fully identifiable because the space of the parameters for the model, $\mathbb{R}^X \times \mathbb{R}^{NX} \times \mathbb{R}^{NT}$ has a higher dimension than that of the model space, $\mathcal{M}$, as we show later. Therefore, the mapping Θ in Equation 3.4 cannot be injective,[5] since we cannot find a one-to-one mapping from a higher dimension space to a lower one. In practice,

[5] A transformation, Θ, which maps set A to set B is injective if $\forall a_1, a_2 \in A,\ \Theta(a_1) = \Theta(a_2) \Leftrightarrow a_1 = a_2$ (which implies that different points get mapped to different points).

this means that we can find transformations of the parameters

$$\{\alpha_x, \beta_x^{(i)}, \kappa_t^{(i)}\} \rightarrow \{\hat{\alpha}_x, \hat{\beta}_x^{(i)}, \hat{\kappa}_t^{(i)}\} \tag{3.5}$$

such that

$$\Theta(\alpha_x, \beta_x^{(i)}, \kappa_t^{(i)}) = \Theta(\hat{\alpha}_x, \hat{\beta}_x^{(i)}, \hat{\kappa}_t^{(i)}) \tag{3.6}$$

We call the transformations of the parameters which satisfy Equation 3.6 "invariant", because the fitted mortality rates do not change when they are applied to the parameters. The additional degrees of freedom in these invariant transformations correspond to the additional dimensions of the parameter space compared with the model space.

Because $\{\alpha_x, \beta_x^{(i)}, \kappa_t^{(i)}\}$ and $\{\hat{\alpha}_x, \hat{\beta}_x^{(i)}, \hat{\kappa}_t^{(i)}\}$ give identical fitted mortality rates and therefore fit observed data equally well, there is no statistical reason to choose between them. In practice, in order to specify a unique set of parameters, constraints independent of the data are imposed - so called "identifiability constraints". This has the effect of reducing the number of degrees of freedom from the number of parameters. Mathematically, imposing constraints restricts the original parameter space, $\mathbb{R}^X \times \mathbb{R}^{NX} \times \mathbb{R}^{NT}$, to a subspace, $\mathcal{P}$, which has fewer dimensions. The aim is to select a subspace, $\mathcal{P}$, which has the same dimension as the model space, $\mathcal{M}$, which allows for a one-to-one mapping between the reduced parameter space and the model space. Reducing the dimension of the parameter space can also be achieved by reparameterising the model in a "maximally invariant" form, as discussed in Appendix 3.B.

It is important to know the number of dimensions of the model space, not only to ensure that our model is uniquely estimated, but also because this value is used to penalise the likelihood or deviance functions in measures of the goodness of fit, such as the Bayes Information Criterion. A failure to correctly determine the number of free parameters in a model may therefore distort tests of the goodness of fit, such as those performed in Cairns et al. (2009) and Haberman and Renshaw (2011), and potentially leads to an incorrect assessment about which model gives a superior fit to data. One specific example of this is discussed in Appendix 3.A.

3.3 Identifiability in the Lee-Carter model

This general lack of identifiability in mortality models has been recognised for a long time. One of the first and most significant AP mortality models was introduced in Lee and Carter (1992) (referred to as the LC model). This has a single age/period term (i.e., $N = 1$ in Equation 3.1) and can be written as

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \tag{3.7}$$

The study of Lee and Carter (1992) was aware that these parameters are not unique as they can be transformed in the following two ways

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \left\{\alpha_x, \frac{1}{a}\beta_x, a\kappa_t\right\} \tag{3.8}$$

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \{\alpha_x - b\beta_x, \beta_x, \kappa_t + b\} \tag{3.9}$$

and the fitted mortality rates will be unchanged. The existence of invariant transformations means that the model possesses identifiability issues, since no one set of parameters is determined uniquely from the data.

We can see that Equation 3.8 implies that the "scales" of β_x and κ_t are unidentified since $\|\beta_x\| \neq \|\hat{\beta}_x\|$ and similarly for κ_t. In addition, we can say that Equation 3.9 implies that the "location" of κ_t is unidentified.[6] The locations and scales of the age and period terms in the LC model therefore have no independent significance, because different sets of parameters, with different locations and scales, will give exactly the same observable quantities, such as fitted mortality rates.

To overcome this lack of identifiability, Lee and Carter (1992) imposed additional constraints on the parameters which are unrelated to the underlying data.[7] As Equations 3.8 and 3.9 have two free parameters, a and b, we require an additional two arbitrary identifiability constraints to uniquely specify the model. Lee and Carter (1992) imposed $\sum_x \beta_x = 1$ and $\sum_t \kappa_t = 0$. These identifiability constraints have subsequently become widely adopted by most model users. A general set of LC parameters (found from the

[6] Scale and location have their intuitive meanings that the "scale" of a set of parameters relates to how spread out they are, whilst "location" refers to their position (i.e., what numerical values they take). More precisely for β_x, we could define the scale of a parameter set as $S = \max(\beta_x) - \min(\beta_x)$ and the location, $L = \frac{\sum_x \beta_x}{XS}$, where X is the number of ages in the range of x, with similar definitions for κ_t. However, these formal definitions provide little by way of additional meaning.

[7] We say that the transformations in Equations 3.8 and 3.9 cause issues with the *identifiability* of the model. *Identification* of the model is accomplished by imposing a set of identifiability constraints and using the invariant transformations to satisfy these constraints.

data via some estimation method) can be transformed into the constrained parameter set using the transformation in Equation 3.8 and choosing $a = \sum_x \beta_x$ and then by using the transformation in Equation 3.9 with $b = -\frac{1}{T}\sum_t \kappa_t$.

We can see that imposing any set of identifiability constraints is achieved by using these transformations with specific values of the free parameters a and b. Intuitively, we might think of the imposition of the identifiability constraints as reducing the number of effective parameters in the LC model. The LC model has $2X + T$ parameters. However, the invariant transformations of the model show that two of these degrees of freedom do not have any impact on the fit to data. Imposing the identifiability constraints involves transforming an arbitrary set of parameters to our chosen set by using the transformations with specific values of these parameters and so can be thought of as "using up" the degrees of freedom in a way that does not affect the fitted mortality rates. We will therefore have a total of $2X + T - 2$ parameters which are determined by the data when fitting the model, and another two which are determined by imposing the identifiability constraints.

In the terminology of Section 3.2.2, the unconstrained parameter space of the LC model has dimension $2X + T$, but the model space, $\mathcal{M}$, has dimension $2X + T - 2$. The identifiability constraints therefore restrict the parameters to the $2X + T - 2$ dimensional subspace, $\mathcal{P}$, of the full parameter space, $\mathbb{R}^X \times \mathbb{R}^{NX} \times \mathbb{R}^{NT}$, allowing for an injective mapping between the restricted parameter space, $\mathcal{P}$, and the model space, $\mathcal{M} \subset \mathbb{R}^{X \times T}$.

We interpret the constraints used in Lee and Carter (1992) as setting first the "normalisation" of β_x in order to identify its scale and second the "level" of κ_t to be centred on zero to identify its location. However, the location and scale chosen still do not possess any independent meaning, since they are wholly dependent upon the identifiability constraints chosen. Because they do not depend upon the data, these additional identifiability constraints are arbitrary. While they might allow us to interpret the parameters in terms of their demographic significance,[8] this interpretation nevertheless depends entirely on the user's judgement, rather than on the underlying data.

For instance, the constraint that $\sum_t \kappa_t = 0$ in the Lee-Carter model allows us to interpret κ_t as representing deviations away from an "average" level of the fitted mortality rates

[8] Demographic significance is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

across the historical period of interest, since it has the consequence that

$$\alpha_x = \frac{1}{T}\sum_t \eta_{x,t} \tag{3.10}$$

The constraint $\sum_t \kappa_t = 0$, therefore, means that α_x can be interpreted as the average mortality rate at each age over the period of the data.[9]

However, the constraint $\kappa_1 = 0$ is just as reasonably imposed in Renshaw and Haberman (2003c), with the interpretation that the period functions represent the falls in mortality from an initial level.[10] Imposing this constraint means that $\alpha_x = \ln(\mu_{x,1})$, i.e., it has the demographic significance that it is the first year of the fitted mortality surface. Accordingly, model users must be careful not to rely on a particular interpretation for the parameters when making mathematical statements about the model or when projecting it. For instance, we should not directly compare values of κ_t for different populations, since different arbitrary identifiability constraints can result in very different estimated values of the parameters.

The use of arbitrary identification constraints has become almost universal amongst users of the LC model. An alternative approach, proposed by Nielsen and Nielsen (2014), is to reparameterise the model to give a set of "maximally invariant" parameters. These will be chosen to avoid any identification issues, but convey the same information and achieve the same fit to data. This approach and its drawbacks are discussed in Appendix 3.B.

3.4 Identifiability in models with non-parametric age functions

We define models with non-parametric age functions in Chapter 2 as those where the values of the age functions $\beta_x^{(i)}$ at different ages x are fitted without any a priori shape

[9] If ordinary least squares is used to estimate the parameters in the model, the estimator for α_x is $\frac{1}{T}\sum_t \ln\left(\frac{d_{x,t}}{E^c_{x,t}}\right)$, i.e., the unweighted average of observed mortality rates. However, this will not be true if other estimation methods are used, where α_x will be a weighted average, where the weights are related to the exposure to risk over the period. Imposing $\frac{1}{T}\sum_t \ln\left(\frac{d_{x,t}}{E^c_{x,t}}\right)$ a priori onto a model will therefore reduce the goodness of fit to the data if alternative fitting procedures are used. The impact of this is discussed further in Appendix 3.A.

[10] This would involve applying the transformation in Equation 3.9 with $b = -\kappa_1$.

across ages. Age is treated as an unknown factor in the model rather than as a regressor with a known form.[11] It is important to recognise that this usage differs from other definitions of "non-parametric" employed in statistics and actuarial science. For the avoidance of doubt, we specifically use the term to refer to whether we assume a specific shape for the age functions in Equation 3.1 a priori.

All AP mortality models with non-parametric age functions are extensions of the LC model, as discussed in Booth et al. (2002) and Renshaw and Haberman (2003b). The number of age/period terms in the model is usually found by maximising the fit to data, whilst their shape can be found through principal component analysis using singular value decomposition, as in Booth et al. (2002), Renshaw and Haberman (2003b), Hatzopoulos and Haberman (2009) and Yang et al. (2010).

We can see from consideration of Equation 3.3 that models with non-parametric age/period terms are not fully identified, since we can transform them using

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} = \{\alpha, \beta A^{-1}, A\kappa\} \tag{3.11}$$

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} = \{\alpha - \beta B, \beta, \kappa + B\mathbf{1}^\top\} \tag{3.12}$$

where A is an $(N \times N)$ matrix whose only constraint is that it needs to be invertible, and B is a $(N \times 1)$ matrix.

Theorem 3.1. *The transformations in Equations 3.11 and 3.12 are the only invariant transformations for the model in Equation 3.3.*

Sketch of Proof Assume, without loss of generality, that the matrix β has full column rank N and κ is of full row rank N. If not, the model is poorly chosen and we could use a model with fewer age/period terms and achieve the same fit to data.

Further, assume that we have two sets of parameters giving the same fitted mortality rates. Then

$$\begin{aligned} \alpha\mathbf{1}^\top + \beta\kappa &= \hat{\alpha}\mathbf{1}^\top + \hat{\beta}\hat{\kappa} \\ \beta\kappa - \hat{\beta}\hat{\kappa} &= (\hat{\alpha} - \alpha)\mathbf{1}^\top \\ &= C\mathbf{1}^\top \end{aligned}$$

[11] For this reason, we could alternatively refer to non-parametric age functions as "factorial" age functions.

for C some arbitrary $(X \times 1)$ matrix. From this, we can multiply both sides by $\hat{\beta}^\top$

$$\hat{\beta}^\top \beta\kappa - \hat{\beta}^\top \hat{\beta}\hat{\kappa} = \hat{\beta}^\top C\mathbf{1}^\top$$

and, as $\hat{\beta}$ is of full column rank, $\hat{\beta}^\top \hat{\beta}$ is invertible and so

$$\hat{\kappa} = (\hat{\beta}^\top \hat{\beta})^{-1}\hat{\beta}^\top \beta\kappa - (\hat{\beta}^\top \hat{\beta})^{-1}\hat{\beta}^\top C\mathbf{1}^\top$$

Defining $A = (\hat{\beta}^\top \hat{\beta})^{-1}\hat{\beta}^\top \beta$ and $B = (\hat{\beta}^\top \hat{\beta})^{-1}\beta^\top C$, we see this is of the same form as the composition of the transformations in Equations 3.11 and 3.12 on κ, with the forms of $\hat{\beta}$ and $\hat{\alpha}$ following directly from this. □

By analogy with the LC model, it should be clear that these transformations represent the generalisation of Equations 3.8 and 3.9 for models with more than one non-parametric age/period term. These are the general invariant transformations of the model. Again, we can see that the existence of these invariant transformations means that the scales and angles of the age and period functions are not identifiable by the model (i.e., not defined by the data), since

$$\|\hat{\beta}_x^{(i)}\| = \|\beta_x^{(i)} A^{-1}\| \neq \|\beta_x^{(i)}\|$$
$$< \hat{\beta}_x^{(i)}, \hat{\beta}_x^{(j)} > = < \beta_x^{(i)} A^{-1}, \beta_x^{(j)} A^{-1} > \neq < \beta_x^{(i)}, \beta_x^{(j)} >$$

i.e., different sets of identifiability constraints will give different scales and angles between the age/period terms. In addition, from Equation 3.12 we see that the locations of the $\kappa_t^{(i)}$'s are unidentified in the same way as in the LC model. Since the scales, angles and locations of the parameters are not defined by the data, we are free to impose them through our choice of identifiability constraints.

This also has consequences for any graphs of the different parameters, with some aspects of any graph not being meaningful, since they depend purely on the arbitrary choice of identifiability constraint. For example, in a graph of $\kappa_t^{(i)}$vs.t, the lack of identifiability in the levels of $\kappa_t^{(i)}$ due to be Equation 3.12 means that the position of the x-axis is not meaningful, since it is just a consequence of an identifiability constraint on the level of $\kappa_t^{(i)}$. Similarly, the scale on the y-axis is not meaningful, since it depends on the normalisation scheme chosen.

By interpreting the angle between different period functions as their correlation, as discussed in Section 3.2, we also see that the lack of identifiability issues in AP mortality

model means that correlations between different period functions are also not meaningful, since they too depend upon the arbitrary identifiability constraints. More generally, the behaviour of any one period function has no objective meaning unless it is also true of any linear combination of all of the period functions. This has important consequences when performing graphical checks on the fitted parameters, and also when we come to project a model, as discussed in Section 3.9.

In the terminology of Section 3.2.2, we see that a general AP model of the form in Equation 3.3 has $X + N(X + T)$ parameters, i.e., the parameter space has dimension $X + N(X + T)$. However, the invariant transformations in Equations 3.11 and 3.12 have $N(N+1)$ parameters which implies that we need to impose $N(N+1)$ identifiability constraints in order to specify a unique set of parameters. This means that the restricted parameter space, $\mathcal{P}$, is an $X + N(X + T) - N(N + 1)$ dimensional subspace of $\mathbb{R}^X \times \mathbb{R}^{NX} \times \mathbb{R}^{NT}$, and, correspondingly, the model space $\mathcal{M}$ is an $X + N(X + T) - N(N + 1)$ dimensional subspace of $\mathbb{R}^{X \times T}$.

The $N(N + 1)$ constraints imposed will still be arbitrary in the sense that they are entirely the choice of the model user. It is impossible to choose between models with the same structure in Equation 3.1 and the same fitting procedure but different identifiability constraints by statistical methods. However, the different terms in them may have different subjective demographic significance depending upon the identifiability constraints imposed.

3.5 Identifiability in the LC2 model

In Section 3.3, we saw how the different identifiability issues were solved in the simplest and most commonly used AP mortality model. We now take the intuition derived from that model and also the theory discussed in Section 3.4 and apply them to the next simplest AP mortality model with non-parametric age functions. The two-term model in Renshaw and Haberman (2003b) (which we shall refer to as the LC2 model) is usually written as

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} \tag{3.13}$$

The LC2 model applies the same normalisation scheme to the age functions to set their scale and the same level for the period functions to set their location as in the original

LC model. Doing so, however, can lead to identifiability issues in this more complicated model as we now show.

3.5.1 Location

Because the location of the period functions is not identifiable, Renshaw and Haberman (2003b) set their level by imposing $\sum_t \kappa_t^{(i)} = 0$ for $i = 1, 2$. As with the LC model, this gives the static age function the demographic significance of representing "average" mortality rates across the period range of the data. This does not cause any additional issues for the LC2 model, so long as it is imposed via an identifiability constraint on κ_t and not by imposing the form of α_x (as discussed in Appendix 3.A).

3.5.2 Scale

To set the scale of the age/period terms, Renshaw and Haberman (2003b) imposed the constraint $\sum_x \beta_x^{(i)} = 1$ for $i = 1, 2$, again, in order to be consistent with the convention established by Lee and Carter (1992). However, the justification for this normalisation scheme makes most sense under the assumption that $\beta_x^{(i)} \geq 0$ for all x - indeed, this is imposed on the LC model in Haberman and Renshaw (2009) at the expense of goodness of fit to the data. If $\beta_x^{(i)} \geq 0$, then $\sum_x \beta_x^{(i)} = 1$ constrains the age function to be in the range $[0, 1]$. The values of $\beta_x^{(i)}$ therefore can be felt to represent a proportion of the factor $\kappa_t^{(i)}$ impacting mortality at age x. In general, however, it may be the case that $\beta_x^{(i)} < 0$ at some ages, especially in models with multiple age/period terms. If so, the interpretation of the age functions as measuring the proportion of the change is no longer applicable.

Figure 3.1 shows the age functions from the LC2 model fitted to data for men in the UK[12] with the constraint $\sum_x \beta_x^{(i)} = 1$ for $i = 1, 2$. We see that if $\beta_x^{(i)} \leq 0$ for some x, as is the case for the second age function, then the identifiability constraint on the age function no longer limits it to a particular range of values. Indeed, $\beta_{x_1}^{(i)}$ can take arbitrarily high values, as long as there exists a correspondingly low $\beta_{x_2}^{(i)}$ to compensate. This is in contrast to $\beta_x^{(1)}$, which is greater than zero for all ages, and hence is comparatively close to zero across the whole age range.[13] This undermines the rationale for selecting

[12] Data for men aged 50 to 100 in the UK from 1950 to 2011 from the Human Mortality Database (Human Mortality Database (2014)).

[13] In Figure 3.1, $0.003 \geq \beta_x^{(1)} \geq 0.024$, while $-1.58 \geq \beta_x^{(2)} \geq 1.46$, i.e., roughly two orders of magnitude difference, with a corresponding impact on the period functions.

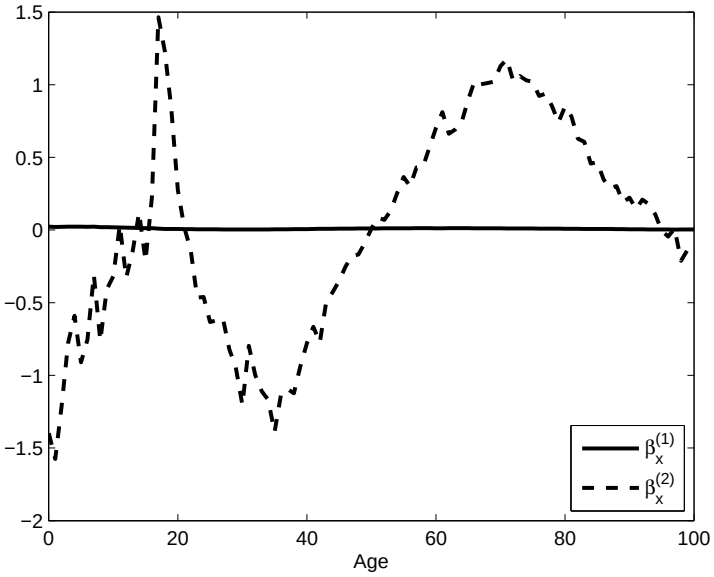


FIGURE 3.1: LC2 age functions with $\sum_x \beta_x^{(i)} = 1$

a common normalisation scheme for the age functions, which was to aid comparisons of the relative importance of the different age/period terms.

The identifiability constraint $\sum_x \beta_x^{(i)} = 1$ can also, theoretically, lead to numerical problems when fitting the model to data. In practice, the constraint is imposed by taking the set of parameters generated by the fitting algorithm (which do not have any identifiability constraints imposed) and using the transformation in Equation 3.8 with $b = \sum_x \beta_x^{(i)}$, i.e., $\hat{\beta}_x^{(i)} = \frac{1}{\sum_\xi \beta_\xi^{(i)}} \beta_x^{(i)}$. This gives an equivalent set of parameters (with the same fit to the data), but where $\sum_x \hat{\beta}_x^{(i)} = 1$ by construction. If, however, $\sum_x \beta_x^{(i)} = 0$ for whatever reason, this procedure will fail as applying the transformation involves dividing by zero, even if the age function fitted originally by the algorithm is reasonable. While this is unlikely, it is far more common that we find $\sum_x \beta_x^{(i)} \approx 0$, which will then lead to the revised parameters (with the constraint imposed) being infeasibly large, and which may, in turn, generate problems with the fitting algorithm.

Both of these problems with the normalisation scheme are caused because simple summation over x is not a true norm. A true norm, $\|v\|$, for a vector space, $\mathcal{V}$, of a vector, v, is defined by the properties

1. $\|v\| \geq 0 \; \forall v \in \mathcal{V}$;
2. $\|v\| = 0 \iff v = 0$;

3. $\|av\| = |a|\|v\| \; \forall a \in \mathbb{R}$; and

4. $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$.

These properties mean that we can use a true norm to define distances and scales within the vector space and therefore make them useful when specifying a normalisation scheme. However, we see that $\sum_x \beta_x^{(i)}$ is not a true norm in $\mathbb{R}^X$, since we can have $\sum_x \beta_x^{(i)} < 0$ and $\sum_x \beta_x^{(i)} = 0$ does not mean that $\beta_x^{(i)} = 0 \; \forall x$. Therefore, we are not able to use this normalisation scheme to compare scales for the age functions, and cannot assume that $\sum_x \beta_x^{(i)} > 0$ in our fitting algorithms when we come to impose the identifiability constraints.

Normalisation schemes using true norms on $\mathbb{R}^X$, such as $\sum_x |\beta_x^{(i)}| = 1$ or $\sum_x (\beta_x^{(i)})^2 = 1$, will not suffer from these issues. When it comes to normalising the fitted age function, a procedure using a true norm for the normalisation scheme will never involve division by zero if the transformation in Equation 3.8 is used with any non-trivial age functions. Therefore, in most circumstances, normalisation schemes based on true norms will be preferable.[14]

However, we note that normalisation schemes based on true norms are not perfectly identified, since the transformation

$$\{\hat{\beta}_x^{(i)}, \hat{\kappa}_t^{(i)}\} = \{-\beta_x^{(i)}, -\kappa_t^{(i)}\} \tag{3.14}$$

is an invariant transformation of the parameters where the new parameters still satisfy the identifiability constraints. In principle, we could solve this by choosing alternative sets of normalisation constraints, for instance

$$sign\left(\sum_x \beta_x^{(i)}\right) \sum_x \left(\beta_x^{(i)}\right)^2 = 1$$

which are still based on using true norms but are not invariant to changing the sign of the age function. However, the specific transformation causing this problem has few practical consequences when fitting the model, since the transformation is not continuous. When fitting the LC or LC2 models using maximum likelihood techniques, for instance, we make small adjustments to the parameters at each iteration and so it is not possible to move smoothly from one set of acceptable parameters to another when

[14] An obvious choice would be a normalisation scheme that is consistent with the standard Euclidean inner product, i.e., the Euclidean norm on $\mathbb{R}^X$, $\|\beta_x^{(i)}\| = \sum_x (\beta_x^{(i)})^2 = 1$. However, this is not essential and an alternative normalisation scheme based on another true norm of $\mathbb{R}^X$ may be preferred if it is more convenient, as it is in Chapter 5.

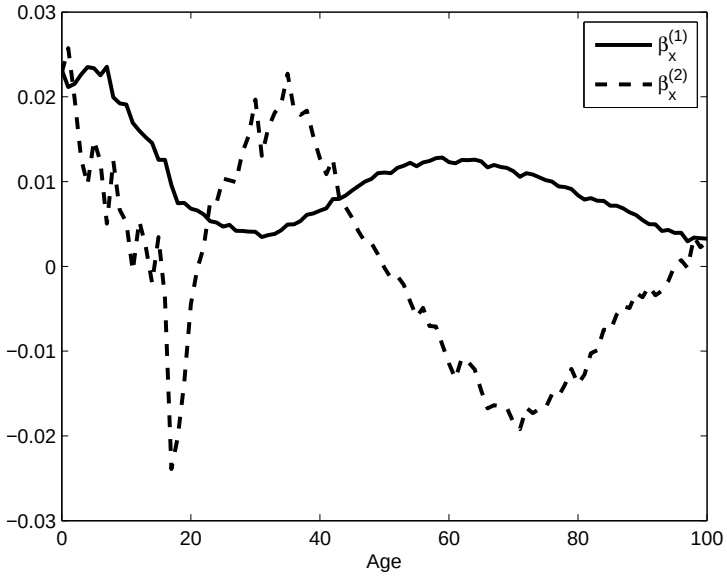


FIGURE 3.2: LC2 age functions with $\sum_x |\beta_x^{(i)}| = 1$

fitting the model. In addition, the transformation in Equation 3.14 can be applied to any set of parameters after fitting the model and, hence, can be used to select the sign of the age function based on the judgement of the user when reviewing the fitted parameters.

To illustrate this, consider the age functions shown in Figure 3.2 which fit the LC2 model to the same data as in Figure 3.1 with the normalisation scheme $\sum_x |\beta_x^{(i)}| = 1$. This normalisation scheme gives a model with exactly the same fit to the data, but the estimated parameters for the age and period functions are now of the same order of magnitude,[15] which may make this model easier to project. We also avoid the possibility of any computational problems when imposing the identifiability constraint, since the divisor, $\sum_x |\beta_x^{(i)}|$, will not be zero for any non-trivial age function.

3.5.3 Rotation

We established in Section 3.4 that $N(N+1)$ constraints were necessary to restrict the parameters in a general AP mortality model with non-parametric age functions, due to the number of free parameters in the transformations in Equations 3.11 and 3.12. In the context of the LC2 model, this means that we would require six identifiability constraints. However, only four identifiability constraints (two on the level of the two period functions, two on the normalisation of the two age functions) were described

[15] In Figure 3.1, $0.003 \geq \beta_x^{(1)} \geq 0.024$, while $-0.024 \geq \beta_x^{(2)} \geq 0.026$, i.e., the same order of magnitude.

in Renshaw and Haberman (2003b). We, therefore, have an additional two invariant transformations of the parameters which give the same fit to data and which satisfy the constraints already explicitly imposed by Renshaw and Haberman (2003b). These can be written as

$$\begin{aligned}\begin{pmatrix}\hat{\beta}_x^{(1)}\\ \hat{\beta}_x^{(2)}\end{pmatrix} &= \begin{pmatrix}\theta & 1-\theta\\ 0 & 1\end{pmatrix}\begin{pmatrix}\beta_x^{(1)}\\ \beta_x^{(2)}\end{pmatrix}\\ \begin{pmatrix}\hat{\kappa}_t^{(1)}\\ \hat{\kappa}_t^{(2)}\end{pmatrix} &= \frac{1}{\theta}\begin{pmatrix}1 & \theta-1\\ 0 & \theta\end{pmatrix}\begin{pmatrix}\kappa_t^{(1)}\\ \kappa_t^{(2)}\end{pmatrix}\end{aligned} \tag{3.15}$$

and

$$\begin{aligned}\begin{pmatrix}\hat{\beta}_x^{(1)}\\ \hat{\beta}_x^{(2)}\end{pmatrix} &= \begin{pmatrix}1 & 0\\ 1-\phi & \phi\end{pmatrix}\begin{pmatrix}\beta_x^{(1)}\\ \beta_x^{(2)}\end{pmatrix}\\ \begin{pmatrix}\hat{\kappa}_t^{(1)}\\ \hat{\kappa}_t^{(2)}\end{pmatrix} &= \frac{1}{\phi}\begin{pmatrix}\phi & 0\\ \phi-1 & 1\end{pmatrix}\begin{pmatrix}\kappa_t^{(1)}\\ \kappa_t^{(2)}\end{pmatrix}\end{aligned} \tag{3.16}$$

These transformations can be thought of as "rotations" of the age/period functions, because they change the angle between age and period functions, but the normalisation scheme $\sum_x \hat{\beta}_x^{(i)} = 1$ still holds.[16] They also clearly illustrate that we have an additional two degrees of freedom, given by the free parameters θ and ϕ, which do not change the fitted mortality rates but which should be used to impose two more identifiability constraints on the model.

This does not necessarily mean that the model in Renshaw and Haberman (2003b) was poorly identified, however. Although the authors did not explicitly acknowledge the existence of these additional identifiability constraints, their use of singular value decomposition to fit the model imposed them implicitly. By taking singular values (or equivalently, principal components), age and period functions are selected so that $\sum_t \kappa_t^{(i)}\kappa_t^{(j)} = 0$ and $\sum_x \beta_x^{(i)}\beta_x^{(j)} = 0$ for $i \neq j$. We call such age and period functions "orthogonal" to each other as the angle between them defined earlier using the standard inner product will be $\frac{\pi}{2}$. This implicit imposition of additional identifiability constraints leads to a fully identified model.

If alternative fitting methods are used, such as maximum likelihood (e.g., in Brouhns et al. (2002a)) or minimal deviance (e.g., in Renshaw and Haberman (2003a)), then these constraints must be imposed explicitly in order to obtain a fully identified model.

[16] In some respects, Equations 3.15 and 3.16 are more similar to shears than rotations. However, we find that thinking of them as rotations with respect to the original set of parameters is conceptually more helpful.

To impose these orthogonality constraints for a general set of LC2 parameters, we would therefore need to solve $\sum_t \hat{\kappa}_t^{(i)} \hat{\kappa}_t^{(j)} = 0$ and $\sum_x \hat{\beta}_x^{(i)} \hat{\beta}_x^{(j)} = 0$ with the transformed parameters defined by Equations 3.15 and 3.16 in order to find θ and ϕ.

We also note the special case where $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (i.e., $\theta = \phi = 1$ when Equations 3.15 and 3.16 are composed), which relates to the transformation

$$\{\hat{\beta}_x^{(1)}, \hat{\kappa}_t^{(1)}, \hat{\beta}_x^{(2)}, \hat{\kappa}_t^{(2)}\} = \{\beta_x^{(2)}, \kappa_t^{(2)}, \beta_x^{(1)}, \kappa_t^{(1)}\} \tag{3.17}$$

This is an invariant transformation of the parameters where the new parameters still satisfy the identifiability constraints. However, it amounts to simply re-labelling the age/period terms and arises because the identifiability constraints are the same for all age/period terms. Similar to the case in Equation 3.14, this situation could, in principle, be solved by using different identifiability constraints for the different age/period terms, for instance

$$\sum_x |\beta_x^{(1)}| = 1$$
$$\sum_x \left(\beta_x^{(2)}\right)^2 = 1$$

which breaks the symmetry between the different age/period terms and, thus, prevents them being relabelled. However, as with Equation 3.14, the transformation in Equation 3.17 has few practical consequences, since it is not continuous and so it is not possible to move smoothly from one set of acceptable parameters to another when fitting the model. Furthermore, using different identifiability constraints for the different age/period terms conflicts with a desire for their scale to be comparable with each other and, hence, we do not believe that this issue is important in practice.

If maximum likelihood methods are used to estimate the parameters in a model, it is useful that these estimators are independent of each other. This helps to give more efficient fitting algorithms for estimation and is also useful when allowing for parameter uncertainty using the technique of Brouhns et al. (2002b) discussed in Section 3.8. Assuming the canonical link function is used as discussed in Chapter 2, the independence of the estimators can be assessed by consideration of the information matrix for the different

parameters

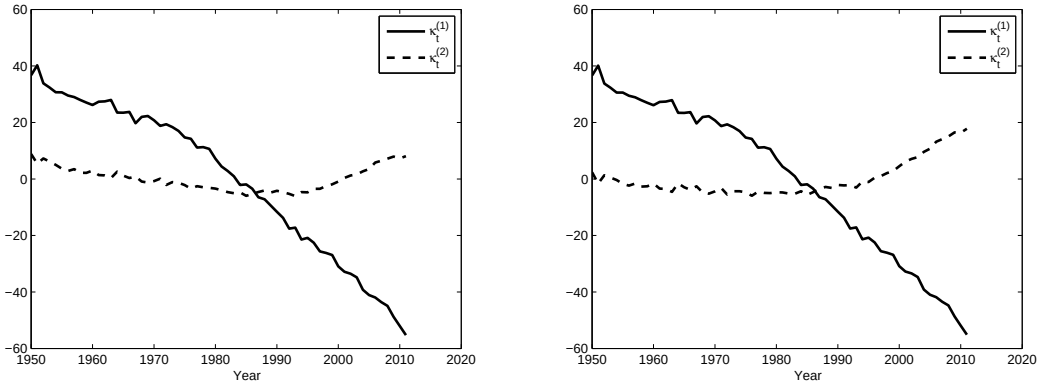
$$
\begin{aligned}
I(\beta_x^{(i)}, \beta_x^{(j)}) &= \mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \beta_x^{(i)} \partial \beta_x^{(j)}}\right] \\
&= -\sum_t \mathbb{V}ar(D_{x,t}) \kappa_t^{(i)} \kappa_t^{(j)} \\
I(\kappa_t^{(i)}, \kappa_t^{(j)}) &= \mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \kappa_t^{(i)} \partial \kappa_t^{(j)}}\right] \\
&= -\sum_x \mathbb{V}ar(D_{x,t}) \beta_x^{(i)} \beta_x^{(j)}
\end{aligned}
$$

Therefore, we see that orthogonal age and period functions are independent of each other if $\mathbb{V}ar(D_{x,t})$ is constant across ages and years.This assumption is implicitly made when using singular value decomposition or principal components analysis to estimate parameters. However, the assumption is not consistent with the use of the Poisson or binomial distribution for death counts, as discussed in Chapter 2. Under these distributions, the variance of death counts depends upon the exposure to risk at different ages, which changes considerably over different ages and years and is more realistic in practice.

In principle, we could impose independent parameter estimates using the transformations in Equations 3.15 and 3.16 with carefully selected values of θ and ϕ to obtain an equivalent set of parameters. Doing so would simply be choosing an alternative (but equally valid) set of identifiability constraints. However, in practice, this would mean constraints that are both more difficult to impose than the traditional orthogonality constraints using the Euclidean inner product, and which lose the connection between the inner product and the sample moments of $\kappa_t^{(i)}$. In practice, imposing $\sum_t \kappa_t^{(i)} \kappa_t^{(j)} = 0$ and $\sum_x \beta_x^{(i)} \beta_x^{(j)} = 0$ for $i \neq j$ to obtain orthogonal age and period functions is a convenient and useful set of identifiability constraints.

Whichever set of constraints is imposed on the angles between different period functions, the most important thing is, however, to impose some form of constraint. A failure to do so may result in the fitting routine failing to converge or, alternatively, the fitting routine may give model parameters which depend upon the initial parameter estimates used in the algorithm. Similarly, the angles between different age functions must also be constrained in order to fully identify the model. This has implications for estimated parameter uncertainty, as discussed in Section 3.8.

We noted in Section 3.2 that the correlation between two different period functions depends on the angle between them. This means that we see that the correlations we



(A) Correlation = 0%

(B) Correlation = −75%

FIGURE 3.3: Period functions from the LC2 model

find between period functions from our fitted parameters depends only on the identifiability constraints chosen, and so are not meaningful. For instance, the constraint $\sum_t \kappa_t^{(i)}\kappa_t^{(j)} = 0$ imposes independence on the period functions over the historical range of the data when they are considered as time series. Figure 3.3 shows period functions for the LC2 model fitted to the same data as above, but with two different constraints on the angles between them. In Figure 3.3a, the period functions are orthogonal whereas, in Figure 3.3b, they have a correlation of -75%.[17] However, both sets of parameters give identical fits to the historical data. This will have important consequences when we come to project the model in Section 3.9.

In situations such as Renshaw and Haberman (2003b), where orthogonality constraints on the age/period terms have been imposed implicitly by the fitting mechanism, we believe that it is important to recognise and state them clearly. Not only will this clarify which features of graphs of the age and period terms are meaningful, it also ensures that we assess the dimension of $\mathcal{P}$ (i.e., the number of degrees of freedom in the model) correctly. This is important when assessing the goodness of fit for the model.

As an example of this, in Haberman and Renshaw (2011), the LC2 model is compared against other mortality models using various measures including the Akaike Information Criterion, Bayes Information Criterion, and Hannan–Quinn Criterion. All of these measures use the number of degrees of freedom (i.e., $dim(\mathcal{P})$) of the model to penalise the log-likelihood. By failing to explicitly state the orthogonality constraints placed on the age/period terms in the LC2 model and, therefore, failing to include them in the count of restrictions placed upon the model parameters, the study overestimates the number

[17] Although the period functions in Figures 3.3a and 3.3b are very similar, the relative large negative correlation is due to the fact that $\kappa_t^{(1)}$ is strongly trending over the period.

of degrees of freedom in the model. This excessively penalises the LC2 model relative to its comparators.

Using the invariant transformations to impose orthogonality on the age and period functions generalises naturally to more complicated models with $N > 2$. Identifiability in-sample in a model with non-parametric age/period terms is therefore not problematic if fitting methods based on singular value decomposition or principal component analysis are used (except for setting the locations of the $\kappa_t^{(i)}$ and the scale for the $\beta_x^{(i)}$ by imposing an appropriate normalisation scheme).

3.6 Identifiability in models with parametric age functions

In contrast to the non-parametric age functions considered above, we define a "parametric" age function to be one which takes a specific functional form that is defined by an algebraic formula, i.e., $\beta_x^{(i)} = f^{(i)}(x;\theta^{(i)})$.[18] In order to specify a mortality model with parametric age functions, we need to define these formulae. Mathematically, AP mortality models with parametric age functions are similar to their non-parametric counterparts, except that the age functions are fixed or selected from a family with a small number of free parameters rather than being allowed to vary freely across $\mathbb{R}^X$. This has important consequences for the identifiability issues in the model.

To illustrate, let us consider the following two pedagogical mortality models

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} \tag{3.18}$$

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + e^{-\lambda x}\kappa_t^{(2)} \tag{3.19}$$

where $\bar{x} = 0.5(X+1)$. The first of these is similar to the widely used Cairns-Blake-Dowd (CBD) model of Cairns et al. (2006a), but with the inclusion of an explicit static age function, and therefore we refer to it as the CBDX model. The second model, which we refer to as the exponential model, uses an exponentially decreasing function of age as the second age function, with the parameter λ being a free parameter of the model determined by the data. Such a model has not been proposed to date, but similar terms have been used within the "general procedure" of Chapter 5.

We say that the formulae used for the age functions in Equations 3.18 and 3.19 "define" these models. Different definitions for the age functions give different models. However,

[18] For this reason, these age functions could also be called "formulaic".

we also define the concept of "equivalence" between models with parametric age functions. Two models are equivalent in this sense if they have different definitions for the age functions, but still give the same fitted mortality rates and hence the same fit to data.

We note that the CBDX model is linear in its parameters, and so can be fitted using generalised linear models, as discussed in McCullagh and Nelder (1983) and Currie (2014). However, since λ is a free parameter of the model, the second age/period term in the exponential model is non-linear in the sense of McCullagh and Nelder (1983, Chapter 11), and so more complicated methods for fitting the model are necessary. Therefore, using parametric age functions is not equivalent to using a linear model except in a few simple cases. We will see below that it is these non-linear cases which tend to have more complicated identifiability issues.

Mathematically, we can see that both models in Equations 3.18 and 3.19 are similar to the LC2 model, but with specific parametric functions for $\beta_x^{(1)}$ and $\beta_x^{(2)}$. One might be tempted to believe that they have exactly the same identifiability issues as those in the LC2 model discussed in Section 3.5. However, the imposition of specific functional forms for the age functions has changed whether the invariant transformations of the LC2 model can be applied in practice.

Because the form of the age functions defines the model being used, these forms cannot change under invariant transformations, otherwise we would obtain a different model. Therefore, we require that any invariant transformations of the model also leave the age functions unchanged, i.e., $\hat{f}^{(i)}(x;\theta^{(i)}) = f^{(i)}(x;\theta^{(i)})$. This restriction reduces the number of invariant transformations, and therefore the number of identifiability constraints which need to be imposed when fitting the model to data. We discuss the implications of this on the different identifiability issues below.

3.6.1 Location

We noted in Section 3.4 that the transformation in Equation 3.12 does not change the form of the age functions. Accordingly, it can still be applied to change the levels of the period parameters in exactly the same manner as described in Section 3.4, whilst leaving the fitted mortality rates and the functional forms of the age functions unchanged. The period functions in models with parametric age functions therefore still have unidentified locations, and so we still need to impose levels on the period parameters in exactly the

same manner as we did in Section 3.5. Most users of such models impose $\sum_t \kappa_t^{(i)} = 0$, consistent with the choice made for models with non-parametric age functions and with a similar interpretation. However, for models which have a specific form of the static age function imposed a priori, this is not necessary, as discussed in Appendix 3.A.

3.6.2 Scale

We see that the transformation in Equation 3.11 takes linear combinations of the old age and period functions in order to create new age/period terms. Therefore, these transformations will change the form of the age functions in a model with parametric age functions. Since the form of the age functions defines the model being used, the transformations in Equation 3.11 cannot be used in models with parametric age functions.

In Section 3.5, we saw that these transformations were useful in models with non-parametric age functions when it came to imposing a normalisation scheme on the age functions and orthogonalising them with respect to each other. This was beneficial as it enabled comparability and near-independence between different age/period terms. It is therefore desirable to also achieve the same properties for models with parametric age functions.

We also see that although using the transformations in Equation 3.11 in models with parametric age functions gives different age functions (and therefore different models), they do not affect the fitted mortality rates: all the models obtained by using these transformations are equivalent in the sense defined above. It therefore makes sense to choose, from the set of models equivalent to the one we are interested in, a model with age functions which have the desirable properties of possessing a standard normalisation scheme and being orthogonal. We discuss how this can be done in this section and Section 3.6.3, respectively.

Most mortality models with parametric age functions have the age functions defined in their simplest and most natural form. However, choosing definitions for their simplicity rather than for desirable statistical properties, such as having a common normalisation scheme, can lead to issues when comparing age and period terms within the same model and between different models. We show this below for the CBDX and exponential models in Equations 3.18 and 3.19, respectively. However, for each of these models we also show how this issue can be resolved by using alternative definitions of the age functions

to give models which have far more comparable age and period terms.

First, let us consider how a common normalisation scheme for the age functions can be achieved in the CBDX model in Equation 3.18. In the LC2 model, Renshaw and Haberman (2003b) imposed the normalisation scheme $\sum_x \beta_x^{(i)} = 1$ on the age functions in the model, using the transformations in Equation 3.11. In contrast, the age functions in Equation 3.18 already have defined scales, i.e., $\sum_x f^{(1)}(x) = \sum_x 1 = X$ and $\sum_x f^{(2)}(x) = \sum_x (x - \bar{x}) = 0$.

However, these defined scales cause problems when it comes to comparing the age/period terms. The most important of these issues is that the scale of $f^{(2)}(x)$ is zero, which is not sensible for a functions which is not identically equal to zero. This is a consequence of using a normalisation scheme which is not based on using a true norm. In Section 3.5, we saw that a more sensible choice of normalisation scheme was to use $\sum_x |\beta_x^{(i)}|$ to define the scales of the age functions. Using this for the CBDX model, we find $\sum_x |f^{(1)}(x)| = \sum_x 1 = X$ and $\sum_x |f^{(2)}(x)| = \sum_x |(x - \bar{x})| = 0.25X^2$ if X is even or $0.25(X-1)(X+1)$ if X is odd.

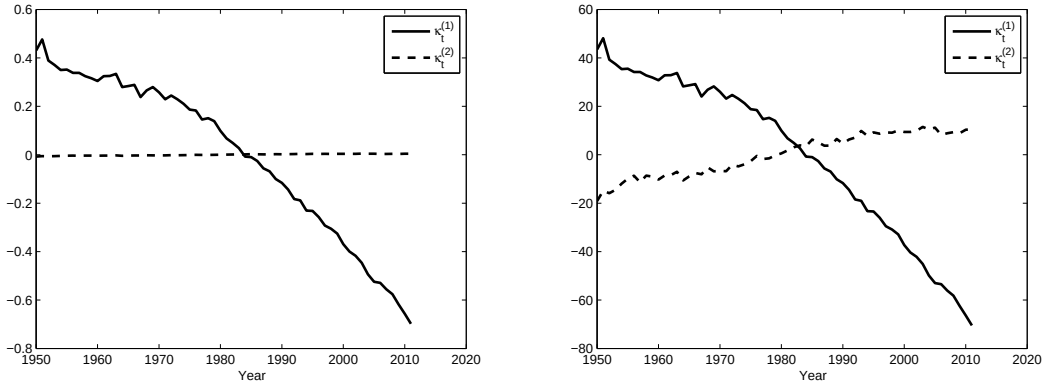
However, this fails to resolve the second problem, which is that different scales are defined for each of the age/period terms, i.e., the scale of the first age function is proportional to the number of ages, X, whilst the scale of the second in proportional to X^2. This makes comparisons difficult, both between the CBDX and LC2 models and between the first and second age/period terms within the CBDX model. The differing scales of the corresponding period functions can also lead to numerical problems when we try to project them using multivariate methods, as discussed in Section 3.9.

To ensure that the age functions have the same scale, we need to define a model equivalent to that in Equation 3.18 where the age functions have this property. Trivially, we see that the model

$$\eta_{x,t} = \alpha_x + \frac{1}{X}\kappa_t^{(1)} + \frac{4(x - \bar{x})}{X^2}\kappa_t^{(2)} \tag{3.20}$$

(assuming X is even) is equivalent to the model in Equation 3.18.[19] All that differs between the models in Equations 3.18 and 3.20 is the precise definition of the age functions, although the age functions in both models have the same functional form (i.e., a constant

[19] We can think of this model being obtained by using the transformation in Equation 3.11 on the model in Equation 3.18 with $A = \begin{pmatrix} X & 0 \\ 0 & \frac{1}{4}X^2 \end{pmatrix}$.



(A) Original definition of age functions $f^{(i)}(x)$ (B) Revised definition of age functions $f^{(i)}(x)$

FIGURE 3.4: Period functions from the CBDX model

and a linear function of age, x). In addition, we see that in the model in Equation 3.20, $\sum_x |f^{(i)}(x)| = 1$ for both age functions. In particular, this has the advantage of greater comparability between the age/period terms.

To illustrate the impact of ensuring that the age functions have a common normalisation scheme, Figure 3.4 shows the period functions from the two CBDX models in Equations 3.18 and 3.20, fitted to the same data as used for the LC2 model in Section 3.5, with both the original and the revised normalisation schemes. We see that the magnitude of the different period functions fitted with the original model in Equation 3.18 differs enormously.[20] This can be a problem as most numerical algorithms for analysing time series are optimised to work best on series of comparable orders of magnitude. In contrast, the revised CBDX model in Equation 3.20 gives period functions of comparable magnitude.[21] The common scale also means that it is easier to compare these period functions with those in Figure 3.3 from the LC2 model.

Turning now to the exponential model in Equation 4.17, we find similar issues for the normalisation scheme of the age functions. In the exponential model, $\sum_x |f^{(1)}(x)| = X$ as before for the CBDX model, which can be dealt with in exactly the same manner. In addition, $\sum_x |f^{(2)}(x;\lambda)| = \sum_x e^{-\lambda x} = \frac{e^{-\lambda}(1-e^{-\lambda(X+1)})}{1-e^{-\lambda}} \approx \frac{e^{-\lambda}}{1-e^{-\lambda}}$ for the second age function. Not only will this be different from the scale of the first age/period term, but the scale is a function of the free parameter λ. Since λ varies during the fitting process, this will alter the scale of $f^{(2)}(x;\lambda)$. Hence, λ will be trying to fulfil two purposes simultaneously: first, describing the shape of the age function and second, determining its scale, i.e., the relative importance of the age/period term. This confusion of different

[20] $-0.70 \geq \kappa_t^{(1)} \leq 0.48$ and $-0.01 \geq \kappa_t^{(2)} \leq 0.05$, i.e. they differ by an order of magnitude.
[21] $-70.5 \geq \kappa_t^{(1)} \leq 48.1$ and $-19.1 \geq \kappa_t^{(2)} \leq 11.5$, i.e., they are the same order of magnitude.

purposes can cause numerical instability in most fitting algorithms, which may be one reason why age functions with free parameters have not been commonly used in practice.

For the CBDX model, we obtained a common normalisation scheme for the age functions by choosing slightly different definitions for the age functions, i.e., we defined alternative age functions which were equal to the original ones, but rescaled by $\sum_x |f^{(i)}(x)|$. For the exponential model we do the same thing, to obtain

$$\eta_{x,t} = \alpha_x + \frac{1}{X}\kappa_t^{(1)} + \frac{1-e^{-\lambda}}{e^{-\lambda}(1-e^{-\lambda(X+1)})}e^{-\lambda x}\kappa_t^{(2)} \tag{3.21}$$

The only difference in this case is that the second age function is rescaled by a function of the free parameter, λ, rather than a constant in the case of the CBDX model. Again, we see that the age functions have the same functional forms (a constant and an exponential function of age) as before, but with the normalisation scheme $\sum_x |f^{(2)}(x;\lambda)| = 1 \ \forall\lambda$ as λ is varied when fitting the model. This contrasts with the model in Equation 3.19, and ensures that both age functions have the same normalisation scheme and so are more comparable.

We call age functions such as the revised $f^{(2)}(x;\lambda)$ in Equation 3.19 "self-normalising", as they have the property that our desired normalisation scheme is imposed automatically for all values of the free parameters in the age function (i.e., $\sum_x |f^{(i)}(x;\theta^{(i)})| = 1 \ \forall\theta^{(i)}$). Self-normalisation is an important and useful property. Most importantly, the common normalisation scheme allows for comparability between different age functions (potentially with very different functional forms) in a model, independent of their shape. Furthermore, by allowing the value of the free parameter to describe the shape of the age function, without impacting the scale of the age/period term, we find that self-normalising age functions are considerably more robust (in the sense of being likely to converge) and stable to small changes in the data. For this reason, the age functions used in the "toolkit" in the Appendix of Chapter 5 are all self-normalising with respect to the normalisation scheme $|f^{(i)}(x;\theta^{(i)})| = 1$.[22] However, the trade-off is that the numerical routines are significantly more complicated to implement and may need to be written specially for the specific circumstances, rather than adapted from "off-the-shelf"

[22] We note that, for many age functions, it is considerably simpler to find and use self-normalisation age functions when using the L1 normalisation scheme, $\sum_x |f^{(i)}(x;\theta^{(i)})| = 1$, than the alternative L2 normalisation scheme, $\sum_x \left(f^{(i)}(x;\theta^{(i)})\right)^2 = 1$. This is why the L1 normalisation scheme was selected for use in the general procedure in Chapter 5.

statistical packages.[23]

In summary, we see that, when the age functions in a mortality model are defined parametrically, a common normalisation scheme for all of them can be achieved by defining the age functions carefully. For more sophisticated age functions involving free parameters estimated from the data, this means defining age functions which are self-normalising, so that the normalisation scheme holds for all values of these parameters as they are varied during the fitting procedure.

3.6.3 Rotation

In Section 3.6.2, we saw that for models with parametric age functions, we could ensure that the age functions had the same normalisation scheme by carefully defining them to have this property when we specified the model. The same is also true if we want our age functions to be orthogonal to each other.

Again, similar to Section 3.6.2, we start from the fact that most mortality models have their age functions defined in the simplest form, such as in Equations 3.18 and 3.19. These simple forms are not, necessarily, orthogonal. However, we can define equivalent models where the age functions are orthogonal. Unlike the case of ensuring a common normalisation scheme, however, we will see that orthogonality between age functions is not always a desirable property and may conflict with other desirable properties, such as the terms in the model having distinct demographic significance. Therefore, the choice of whether to define orthogonal age functions or not will depend upon the model in question and the aims of the model user.

For example, consider the CBDX model of Equation 3.18 before normalisation. The model already has orthogonal age functions, since $\sum_x f^{(1)}(x)f^{(2)}(x) = \sum_x (x - \bar{x}) = 0$. However, we could also consider an equivalent model, with simpler definitions of the age functions of the form

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + x\kappa_t^{(2)} \tag{3.22}$$

[23] In practice, there are many age functions where $\sum_x |f^{(i)}(x;\theta^{(i)})|$ cannot be found in closed form, but can be approximated by $\int |f^{(i)}(x;\theta^{(i)})|dx$. In such circumstances, improvements in the stability of the numerical optimisation routine can still be found through approximate normalisation by setting $\hat{f}^{(i)}(x;\theta^{(i)}) = \frac{f^{(i)}(x;\theta^{(i)})}{\int |f^{(i)}(x;\theta^{(i)})|dx}$ and then imposing $\sum_x |f^{(i)}(x;\theta^{(i)})| = 1$ again directly using Equation 3.8 with $a = \frac{1}{\sum_x |f^{(i)}(x;\theta^{(i)})|}$.

This model is more similar to the form of the original CBD model proposed in Cairns et al. (2006a). However, we observe that the age functions are not orthogonal, i.e., $\sum_x f^{(1)}(x)f^{(2)}(x) = \sum_x x = \frac{1}{2}X(X+1)$. It is easy to see that models in Equations 3.18 and 3.22 are equivalent, in that they give the same fitted mortality rates and are linked through a transformation of the form in Equation 3.11. The form of the age functions in Equation 3.18 was introduced in Cairns et al. (2009) and, in practice, has proved far more popular than the simpler age functions in Equation 3.22, in part because it is more robust to fit to data due to the parameter estimates for the period functions being nearly independent of each other. Consequently, we see that defining orthogonal age functions can be desirable, even if it comes at the expense of a slightly more complicated definition of the age functions.

The age functions in the CBDX model are of constant and linear form, i.e., polynomials of order zero and one, respectively. Defining orthogonal age functions, as in Equation 3.18, has not changed this form, merely selected the first two members of the orthogonal family of polynomials, i.e., the Legendre polynomials.[24] The orthogonal age functions in Equation 3.18 have the same demographic significance as the simpler age functions in Equation 3.22, but the additional desirable property of orthogonality. Generalising this, we see that choosing orthogonal age functions does not change their form and hence does not affect their demographic significance when the age functions come from the same functional family (e.g., polynomials).

However, this is not the case when the age functions come from different functional families. We see this by considering the exponential model once more. To define orthogonal age functions for this model, we could select a model equivalent to that in Equation 3.19 with orthogonal age functions, namely

$$f^{(2)}(x;\lambda) = e^{-\lambda x} - \frac{e^{-\lambda}(1-e^{-\lambda(X+1)})}{1-e^{-\lambda}}$$

We see that the age functions in this model are orthogonal as $\sum_x f^{(1)}(x)f^{(2)}(x;\lambda) = 0\ \forall\lambda$. This revised model is equivalent to that in Equation 3.19, as it gives the same fitted mortality rates and the two models are linked by a transformation of the form of that in Equation 3.11.

[24] The Legendre polynomials have a long pedigree, first in mathematical physics, but more recently in the graduation of mortality rates (for instance in Renshaw et al. (1996) and Sithole et al. (2000)). We also note that the third (quadratic) Legendre polynomial is used as an age function in one of the extensions to the CBD model in Cairns et al. (2009).

However, it is likely that we originally selected an exponential function for its demographic significance (e.g., a mortality effect which decreases rapidly with age, such as that associated with the relatively high rate of infant mortality). The redefined $f^{(2)}(x;\lambda)$ will not possess this demographic significance, as it will start positive and then tend rapidly to a negative constant. This lack of demographic significance is unlikely to be desirable. Therefore, orthogonal age functions can conflict with a desire for each age/period term to have distinct demographic significance for models with parametric age functions coming from different functional families.

In summary, we find that orthogonality between age functions makes most sense when the age functions come from the same family, such as polynomials, and therefore can be orthogonalised easily. For models with very different functional forms for the age functions, orthogonalisation is unlikely to be desirable as it will conflict with a desire to give each age/period term distinct demographic significance.

3.7 Identifiability in mixed models

Some AP mortality models have mixed parametric and non-parametric age functions, such as the model of Wilmoth (1990) (excluding the cohort term) and the models used to explore the data in Chapter 5. Other studies, such as Reichmuth and Sarferaz (2008), have proposed extending the LC model with exogenous variables, such as economic or health indicators, which take the form of period functions with a prescribed form. The identifiability issues in such mixed models, however, are similar to those addressed in Sections 3.5 and 3.4 above.

As with models with purely parametric age functions, in mixed models, the prescribed form of the age or period functions means that we must restrict the transformations in Equations 3.12 and 3.11 so that they remain unchanged. For instance, consider the model

$$\eta_{x,t} = \alpha_x + f(x)\kappa_t^{(1)} + \beta_x\kappa_t^{(2)} \tag{3.23}$$

This model has one parametric age function, $f(x)$, and one non-parametric age function, β_x, while the two period functions are freely varying. We see that the transformation in Equation 3.12 is still applicable, as it will not change the form of $f(x)$ and therefore

we still need to define the location of the period functions via an identifiability constraint.

However, we see that the transformation

$$\{\hat{f}(x), \hat{\kappa}_t^{(1)}, \hat{\beta}_x, \hat{\kappa}_t^{(2)}\} = \left\{ f(x), \kappa_t^{(1)} + ab\kappa_t^{(2)}, \frac{1}{a}\beta_x - bf(x), a\kappa_t^{(2)} \right\} \tag{3.24}$$

is an invariant transformation of the model in Equation 3.23 and avoids changing the form of $f(x)$. This is a special case of the general transformation in Equation 3.11, with the matrix, A, taking the restricted form $A = \begin{pmatrix} 1 & ab \\ 0 & a \end{pmatrix}$. We can see that this transformation corresponds to a reduced set of invariant transformations compared with the LC2 model, since it only has two degrees of freedom, compared with the four in the unrestricted matrix, A.

The form of the restrictions on A means that only the scale of β_x (set by a) and the angle between β_x and $f(x)$ (set by b) are undefined. In such a model, it therefore makes sense to impose a standard normalisation scheme on β_x, for example, $\sum_x |\beta_x| = 1$, and an orthogonality constraint between β_x and $f(x)$, i.e., $\sum_x \beta_x f(x) = 0$.

Next, consider the alternative model

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)} K(t) + \beta_x^{(2)} \kappa_t \tag{3.25}$$

where $K(t)$ is either a deterministic function, such as in Callot et al. (2014), or an exogenous variable such as real GDP or an indicator variable to account for an epidemic, such as in Liu and Li (2015), or a war. We also note that this type of model is common in multi-population models where the period function in one population is required to be the same as that in another, for instance, those of Carter and Lee (1992) and Li and Lee (2005). In this case, we see that we can no longer use the unrestricted transformation in Equation 3.12, since the location of $K(t)$ is set a priori. Therefore, we only need to impose a constraint on the level of the remaining period function, such as $\sum_t \kappa_t = 0$.

As with the model in Equation 3.23, we also have a restricted set of transformations of the form in Equation 3.11 in order to avoid changing $K(t)$ in the transformation. In this case, the transformation of the parameters is

$$\{\hat{\beta}_x^{(1)}, \hat{K}(t), \hat{\beta}_x^{(2)}, \hat{\kappa}_t\} = \left\{ \beta_x^{(1)} + \frac{b}{a}\beta_x^{(2)}, K(t), \frac{1}{a}\beta_x^{(2)}, a\kappa_t - bK(t) \right\} \tag{3.26}$$

which leaves $K(t)$ unchanged. In this case, the restricted form of the matrix, A, in Equation 3.11 is $A = \begin{pmatrix} 1 & 0 \\ -b & a \end{pmatrix}$, which can be compared to the restricted form for the model in Equation 3.23.

Similarly, these restricted transformations mean that only the scale of $\beta_x^{(2)}$ (set by a) and the angle between $K(t)$ and κ_t (set by b) are undefined. Consequently, this transformation can be used to impose a normalisation scheme on $\beta_x^{(2)}$ and orthogonalise $K(t)$ and κ_t by means of additional identifiability constraints. In this case, the orthogonalisation of the period functions has the clear interpretation that κ_t explains that part of the variation that is independent of the factor $K(t)$. However, this was not done in Liu and Li (2015), which, in the context of that study, made it difficult to interpret the meaning of κ_t for years when there was an epidemic.

Hence, we see that mixed models act to impose restrictions on the more general set of invariant transformations present in a model with fully non-parametric age functions. These restrictions are specific to different models, and depend upon the specification of the model in question. This is especially common in many multi-population mortality models, such as some of those discussed in Villegas and Haberman (2014), which can be interpreted as mixed models where the form of different age and period functions is common to different populations and hence restricted. Consequently, we must analyse each individual model in order to determine which identifiability issues it possesses and, hence, a suitable set of identifiability constraints to impose.

3.8 Parameter uncertainty and hypothesis testing

3.8.1 Parameter uncertainty

Having obtained a set of parameters by fitting a model to data with some set of arbitrary identifiability constraints, it is common to investigate the degree of uncertainty associated with these estimated parameters. A number of techniques have been developed to do this, for instance

- using the asymptotic normality of parameters estimated by maximum likelihood methods, as in Brouhns et al. (2002b);
- using a "semi-parametric" bootstrap based on Poisson (or binomial) death counts, as in Brouhns et al. (2005);

- using a residual bootstrapping method, such as that developed in Koissi et al. (2006) or the more complicated techniques discussed in D'Amato et al. (2011) and Debón et al. (2008, 2010), and
- using Bayesian Markov chain Monte Carlo (MCMC) methods, as in Czado et al. (2005).

All of these techniques were developed for the LC model, as the simplest and most widely used mortality model. In the following section, we follow this convention and implicitly assume that we are dealing with the LC model. However, in principle, they could all be used with any other AP mortality model.

The first three of these methods have been tested and compared in Renshaw and Haberman (2008) and all four were compared in Li (2014). It is important that any conclusions drawn from them do not depend upon the arbitrary identifiability constraints imposed in the model. Since the fitted mortality rates do not change under the invariant transformations of the model, their variability due to parameter uncertainty should not depend on the identifiability constraints imposed either. Appropriate methods for determining parameter uncertainty should ensure this. Two users of a mortality model, using the same data and method for investigating parameter uncertainty, but using different (but equally valid) identifiability constraints should find the same degree of variability of mortality rates under parameter uncertainty.

It is therefore desirable to start from the difference between the observed and fitted mortality rates, since this will be independent of the identifiability constraints chosen from them model and ensure that our results are consistent with observations. For instance, in Brouhns et al. (2005), Poisson-distributed random death counts were generated at each age and year.[25] The distribution of the bootstrapped death counts is therefore unaffected by which identifiability constraints are imposed. Likewise, the fitting residuals used in Koissi et al. (2006) depend only on the actual and fitted death counts and thus not on the identifiability constraints used in fitting the model. Therefore, estimates of the impact of parameter uncertainty on observable quantities, such as fitted mortality rates or life expectancies, will be independent of the arbitrary identifiability constraints.

[25] In Brouhns et al. (2005), it was assumed that $D_{x,t} \sim Po(d_{x,t})$, i.e., the random death counts follow a Poisson distribution with mean equal to the observed death count. This was modified in Renshaw and Haberman (2008) to $D_{x,t} \sim Po(E^c_{x,t}\mu_{x,t})$, i.e., mean equal to the fitted death counts, which is more consistent with other bootstrapping techniques.

However, estimates for the variability of the model parameters will still only be valid conditional on the chosen set of identifiability constraints. For instance, imposing the constraint $\kappa_1^{(i)} = 0$ in a model will mean that $\kappa_1^{(i)}$ will trivially not show any variability using the Brouhns et al. (2005) or Koissi et al. (2006) methods, but this will not be the case for other choices of constraints. Therefore, the observed parameter uncertainty should be seen only in the context of the identifiability constraints applied.

It is also important to ensure that the model is fully identified when using these bootstrapping approaches. If the model is not fully identified, we may observe spurious variation in the parameters which does not lead to real variability in the fitted mortality rates. This is of most practical relevance with the orthogonality constraints for models such as the LC2 model in Equation 3.13, as these are often overlooked if maximum likelihood or minimum deviance techniques are used to fit the model.

The alternative approach to starting from the difference between observed and expected mortality rates is to consider the distribution of the model parameters directly. However, methods which generate new samples of parameters directly, such as the asymptotic method of Brouhns et al. (2002b) or the Bayesian techniques of Czado et al. (2005), must be used with considerably more care.

First, consider the asymptotic method of Brouhns et al. (2002b). This assumes that the variation of the maximum likelihood parameters is given by the information matrix (i.e., the second derivative of the log-likelihood, $\mathcal{L}$) with respect to the model parameters evaluated at the selected parameter estimates). The first thing to note here is that, in order to identify the model, the likelihood being maximised is the constrained likelihood. Starting from the forms of the likelihood function in Chapter 2, this means that we use Lagrangian multipliers to impose the constraints. For example, to impose the Lee and Carter (1992) model constraints involves adjusting the likelihood function by

$$\mathcal{L}(d_{x,t}; \{\alpha, \beta, \kappa\}) \to \mathcal{L}(d_{x,t}; \{\alpha, \beta, \kappa\}) - \lambda_1 \sum_t \kappa_t - \lambda_2 \left(1 - \sum_x \beta_x\right)$$

Therefore, the information matrix is explicitly dependent upon the identifiability constraints imposed. For instance, we can see this by considering the second derivative of the likelihood with respect to the age function β_x

$$\frac{\partial^2 \mathcal{L}}{\partial (\beta_x)^2} = -\sum_t \mathbb{V}ar(D_{x,t})(\kappa_t)^2$$

if we use the canonical link function, as discussed in Chapter 2. If we apply the transformation in Equation 3.9, β_x is unchanged. However, we have

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial(\hat{\beta}_x)^2} &= -\sum_t \mathbb{V}ar(D_{x,t})(\hat{\kappa}_t)^2 \\ &= -\sum_t \mathbb{V}ar(D_{x,t})(\kappa_t + b)^2 \\ &= \frac{\partial^2 \mathcal{L}}{\partial(\beta_x)^2} + 2b\sum_t \mathbb{V}ar(D_{x,t})\kappa_t - b^2\sum_t \mathbb{V}ar(D_{x,t})\end{aligned}$$

In this case, the form of the information matrix with respect to β_x has changed under a transformation which did not change β_x itself. This needs to be taken into consideration carefully, and may explain the variation in the uncertainty in the fitted mortality rates observed in Renshaw and Haberman (2008) when the identifiability constraints are altered.

Next, we consider Bayesian techniques, such as MCMC. As discussed in Nielsen and Nielsen (2014), these can often appear to solve identifiability issues but in fact confuse and disguise them. The use of Bayesian methods often involves consideration of the posterior distribution, π, of the parameters given by

$$\ln(\pi(\{\alpha, \beta, \kappa\})) = \mathcal{L}(d_{x,t}; \{\alpha, \beta, \kappa\}) + \ln(\phi(\{\alpha, \beta, \kappa\})) + \text{constant}$$

where ϕ is the prior distribution for the parameters. The log-likelihood function, $\mathcal{L}(d_{x,t}; \{\alpha, \beta, \kappa\})$, is unchanged by the invariant transformations of the model parameters and so does not depend upon the chosen identifiability constraints. However, in general, the prior distribution ϕ will change under these transformations, unless it is very carefully chosen. This, in turn, means that the posterior distribution will also vary under the invariant transformations of the model, and so will depend implicitly on any identifiability constraints imposed.

A poorly chosen set of priors implicitly imposes a set of identifiability constraints upon the model. For example, a prior distribution that assumes $\kappa_t^{(i)}$ follows an AR(1) process around zero implicitly imposes a level on the period parameters. These implicit constraints may conflict with the explicit constraints subsequently imposed (such as a subsequent choice of the level of $\kappa_t^{(i)}$). Even when there are no conflicts, this implicit selection of identifiability constraints is opaque and it is not clear which features of the posterior distribution are meaningful and which are mere artefacts of the identifiability

scheme implicit in the prior.

We therefore recommend that the prior distribution of the model parameters, ϕ, is selected so that it is unchanged by the invariant transformations of the model. This enables a single set of identifiability constraints to be imposed upon the model without internal conflicts, with these constraints being clear and transparent to all other model users, and with the posterior distribution being independent of the arbitrary choice of identifiability constraints (just as the likelihood is).

3.8.2 Hypothesis testing

Identifiability issues also have important consequences if hypothesis testing on the parameters is performed. In general, hypotheses cannot be tested on the parameter values directly, since they depend upon the identifiability constraints. For instance, testing the hypothesis $\kappa_T = 0$ in the LC model is meaningless, since we can impose $\kappa_T = 0$ (or any other value) by our choice of identifiability constraint. We might be tempted to find combinations of the parameters which are invariant to the transformations of the parameters and test hypotheses based on these. For instance, we may wish to test the hypothesis that mortality is declining faster at age x_1 than at age x_2 using the LC model. To do this, we might note that the expected value of $B \equiv \frac{\beta_{x_1}}{\beta_{x_2}}$ is invariant under the transformation in Equation 3.11 and so does not depend on the identifiability constraints, making it a suitable candidate for hypothesis testing. However, we would have to take care when using a statistic such as this, since it will be undefined in the case $\beta_{x_2} = 0$, which could not be known before the model is fitted to data. In general, therefore, any tests of hypotheses should be performed on observable quantities such as the fitted mortality rates rather than the model parameters.

Direct hypothesis testing of the parameters in an AP model is not often performed in the literature, and therefore this discussion may appear to be of theoretical interest only. However, it is common to use a variety of statistical tests when determining the time series properties of the period functions. For instance, in Lee and Carter (1992) and Cairns et al. (2011a), Box-Jenkins methods were used to determine the preferred time series process for the period functions of different models. Based on the conclusions above, in many cases, the results of these statistical tests will depend on arbitrary choices made when identifying the model. The properties typically tested, such as stationarity, lagged dependence and cross correlation, will affect our projected mortality rates and

so are matters of great practical importance. We should therefore treat with extreme caution the results of any such analysis. This subject is dealt with further in Section 3.9.

In summary, not only do our estimates of the parameters of an AP model depend on the identifiability constraints when fitting the model, so do our estimates of the uncertainty attached to those parameter estimates. We should therefore avoid testing hypotheses on these parameter estimates, as our results will be dependent on the arbitrary identification scheme imposed. In general, methods of estimating parameter uncertainty which use bootstrapping techniques on the fitted mortality rates, which are independent of our choice of identifiability constraints, are likely to be preferred over methods which target the parameters directly. We must still ensure, however, that our models are fully identified when testing parameter uncertainty, as the parameters in a poorly identified model may show spurious differences in ways which do not affect the variability of the fitted mortality rates.

3.9 Projection

In the preceding sections, we have seen that AP mortality models are not uniquely identified and that we need to impose arbitrary identifiability constraints on the parameters in order fit them to historical data. Two different modellers using the same data and the same model but different arbitrary identification constraints will, consequently, obtain different sets of parameters, but these will give identical fitted mortality rates and, therefore, fits to the data.

For the majority of practical purposes, we not only need to fit a mortality model to historical data but also to use it to project mortality rates into the future. In order to make projections of future mortality rates, we typically model the period parameters as being generated by time series processes and use these to project the parameters stochastically into the future. However, the time series processes generating the period parameters are unknown. To find which processes to use, we typically analyse the fitted parameters by statistical methods, such as the Box-Jenkins procedure, to determine which processes from the ARIMA family provide the best fit.

Nevertheless, when it comes to projecting mortality rates, we need to recognise that there is a fundamental symmetry between the processes of estimating a model and projecting it. The former takes observations to calibrate the model, whilst the latter uses

this calibration to produce projected observations of the future. Due to this symmetry, identification issues which exist when fitting the model may also yield problems when projecting it.

We formalise this by saying that:

> Two sets of model parameters, which give identical fitted mortality rates for the past, should give identical projected mortality rates when projected into the future.

We say that time series processes which satisfy this property are "well-identified".

In particular, the invariant transformations of the parameters of the model which leave the fitted mortality rates unchanged should also leave the projected mortality rates unchanged and, hence, the time series processes used to generate the projected mortality rates unchanged. Consequently, we should use the same time series processes for all sets of parameters from a model which give the same fitted mortality rates. If this is not the case, different processes will be used for different arbitrary identifiability constraints, giving different projected mortality rates. A well-identified time series process should be equally appropriate for all equivalent sets of parameters. For example, we should use the same time series processes to project the period parameters shown in Figure 3.3a for the LC2 model as those shown in Figure 3.3b. Similarly, we should use the time series processes to project the period parameters in the CBDX models in Equations 3.18, 3.20 and 3.22, since all three of these models are equivalent. To confirm this, we need to check that applying the invariant transformations to the parameters, which leave the fitted mortality rates unchanged, do not also affect the time series processes used to project the parameters.

Current practice is to:

1. fit the chosen model to data, imposing any arbitrary identifiability constraints needed to specify the parameters uniquely;
2. select time series processes for projecting the parameters based on either using a statistical method (such as the Box-Jenkins procedure to select the preferred processes from the ARIMA class of models) or by directly choosing the time series

processes to ensure biologically reasonable[26] projections by making an appeal to the demographic significance of the parameters.

However, such an approach often leads to projections of mortality rates which are not well-identified. This is because the second step in the process assumes that the parameters found at the first step are known, rather than merely estimated up to an arbitrary identifiability constraint. This means that current practice builds the arbitrary identifiability constraint into the projection process, ensuring that the projected mortality rates are also arbitrary.

In order to obtain well-identified projections, we need to select our projection methods carefully. This means that the time series model we estimate based on the fitted parameters and project into the future should not change form under the transformations in Equations 3.11 and 3.12. However, we saw in Section 3.4 that we cannot use the transformation in Equation 3.11 in models with non-parametric age functions. Therefore our selection of well-identified projection methods in such models has to be subtly different, as discussed below.

3.9.1 Models with non-parametric age functions

Consider the case of projecting an AP mortality model with non-parametric age functions, which has been fitted using data over the period $[1, T]$ to give mortality rates at time $\tau > T$. From Equation 3.2, we could write this as

$$\eta_{x,\tau} = \alpha_x + \boldsymbol{\beta}_x^\top \boldsymbol{\kappa}_\tau$$

We can also see that the projected mortality rates for the future are unchanged by the use of the invariant transformations of the parameters in Equations 3.12 and 3.11, just as the fitted mortality rates were for the past, i.e.,

$$\eta_{x,\tau} = \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top \hat{\boldsymbol{\kappa}}_\tau$$

[26] The concept of biological reasonableness was introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

where

$$\begin{aligned}\hat{\boldsymbol{\kappa}}_\tau &= A\boldsymbol{\kappa}_\tau + B \\ \hat{\boldsymbol{\beta}}_x^\top &= \boldsymbol{\beta}_x^\top A^{-1} \\ \hat{\alpha}_x &= \alpha_x - \boldsymbol{\beta}_x^\top A^{-1} B\end{aligned}$$

Unlike the fitted parameters, however, the projected $\boldsymbol{\kappa}_\tau$ will be some random variable, whose distribution is a function of the fitted parameters, i.e., $\boldsymbol{\kappa}_\tau = P_\kappa(\tau; \{\boldsymbol{\kappa}\})$. We said previously that we should use the same method of projection for all sets of parameters as a first step in ensuring that the projected mortality rates do not depend upon the identifiability constraints. However, for different identifiability constraints, these processes will be estimated from different sets of fitted parameters, e.g., if we use $P_\kappa(\tau; \{\boldsymbol{\kappa}\})$ to project the untransformed period parameters, we must use $P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\})$ to project the transformed period parameters. If we combine this with the invariance of the projected mortality rates, we have

$$\begin{aligned}\alpha_x + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa}\}) &= \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\}) \\ \alpha_x + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa}\}) &= \alpha_x - \boldsymbol{\beta}_x^\top A^{-1} B + \boldsymbol{\beta}_x^\top A^{-1} P_\kappa(\tau; \{A\boldsymbol{\kappa} + B\}) \\ \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa}\}) &= \boldsymbol{\beta}_x^\top A^{-1} \left[P_\kappa(\tau; \{A\boldsymbol{\kappa} + B\}) - B\right] \\ P_\kappa(\tau; \{\boldsymbol{\kappa}\}) &= A^{-1} \left[P_\kappa(\tau; \{A\boldsymbol{\kappa} + B\}) - B\right] \\ P_\kappa(\tau; \{A\boldsymbol{\kappa} + B\}) &= A P_\kappa(\tau; \{\boldsymbol{\kappa}\}) + B \end{aligned} \tag{3.27}$$

for general $\boldsymbol{\beta}_x$, i.e., that the time series processes we use to project the period functions are location and scale preserving. This is also discussed in Nielsen and Nielsen (2014).

One common practice is to use univariate time series processes to project the period functions, on the grounds that they are uncorrelated over the historical sample. For example, in Hyndman and Ullah (2007, p. 4948), when considering the selection of suitable time series processes for projecting a model with non-parametric age functions, it was stated[27]

> For $N > 1$ this is a multivariate time series problem. However, because of the way the basis functions $\beta_x^{(i)}$ have been chosen, the coefficients $\kappa_t^{(i)}$ and $\kappa_t^{(j)}$ are uncorrelated for $i \neq j$. Therefore it is likely that univariate methods will be adequate for forecasting each series $\kappa_t^{(i)}$, for $i = 1, \ldots, N$.

[27] Notation has been adjusted to reflect that used in the current study.

This logic was reiterated in Hyndman et al. (2013) for a related model, as *"There is no need to consider vector models because the $\kappa_t^{(i)}$ coefficients are all uncorrelated by construction"*.

However, we saw in Section 3.5 that the lack of correlation between the different period functions is a product of the choice of identifiability constraints, and that we could find alternative parameters which gave identical fitted mortality rates which had non-zero correlation. Choosing univariate time series processes will therefore not give well-identified projections, but instead will give projected mortality rates which are dependent upon the identifiability constraints chosen.

The first conclusion we can draw is that we should always use multivariate processes to project mortality models with more than one age/period term. Using a multivariate framework allows us to consider the period functions together and so encourages a unified approach to modelling them, rather than focusing on each period function separately. It also allows the invariant transformations in Equations 3.11 and 3.12 to be applied to the time series processes directly to check whether they are well-identified.

The use of multivariate processes means that the order of integration of each of the time series processes should be the same. We should only consider the stationarity of the vector process as a whole, rather than of its individual components. It is common practice to use the highest order of integration for any of the individual period functions (usually first order) as the order of integration for all of them to avoid identification issues.

We can see this by taking a general multivariate time series process for $\boldsymbol{\kappa}_t$ from the class of VARIMA(p,d,q) processes

$$\Delta^d \boldsymbol{\kappa}_t = \boldsymbol{\mu} + \sum_{s=1}^{p} \Phi_s \Delta^d \boldsymbol{\kappa}_{t-s} + \sum_{r=0}^{q} \Psi_r \boldsymbol{\epsilon}_{t-r} \tag{3.28}$$

and applying the transformations in Equation 3.11 and 3.12 to give

$$\begin{aligned}\Delta^d \hat{\boldsymbol{\kappa}}_t &= A\boldsymbol{\mu} + \Delta^d B - \sum_{s=1}^{p} A\Phi_s A^{-1} \Delta^d B + \sum_{s=1}^{p} A\Phi_s A^{-1} \Delta^d \hat{\boldsymbol{\kappa}}_{t-s} + \sum_{r=0}^{q} A\Psi_r \boldsymbol{\epsilon}_{t-r} \\ &= \hat{\boldsymbol{\mu}} + \sum_{s=1}^{p} \hat{\Phi}_s \Delta^d \hat{\boldsymbol{\kappa}}_{t-s} + \sum_{r=0}^{q} \hat{\Psi}_r \hat{\boldsymbol{\epsilon}}_{t-r}\end{aligned}$$

We therefore see that all general VARIMA(p,d,q) processes are location and scale invariant in the sense of Equation 3.27, and so are well-identified.

However, we also see from this that any specific structure we impose a priori on $\boldsymbol{\mu}$, Φ_s and Ψ_r will not be invariant under these transformations. Our second conclusion is, therefore, that we should not assume any pre-specified locations, scales or correlations between our period functions by assuming a prior structure for the matrices governing the time series processes that drives them.

In practice, in order to be invariant to transformations of the form in Equation 3.11, we should always allow for the possibility of both cross-lags between the time series and contemporaneous correlations between the innovations, even if these are not evident from inspection of the fitted time series. In situations where our arbitrary identification constraints set some of these time series parameters to zero, this will emerge naturally from their estimation and do not need to be imposed by the model user.

Finally, we observe that all VARIMA time series models are invariant to simple rescalings of the period functions, i.e., using the transformation in Equation 3.11, the matrix A being diagonal. Therefore, all time series processes are invariant under alternative choices of normalisation scheme. However, having a consistent scale for all period functions is desirable as it assists with the numerical estimation of the time series parameters.

In summary, the use of multivariate time series processes means that we should not treat the period functions differently when projecting them, as the invariant transformation in Equation 3.11 means that the age/period terms are interchangeable, which, in turn, means that we can rotate them without changing the fit to data or the demographic significance of any of the parameters.

3.9.2 Projecting the LC2 model

As a practical example of this, consider projecting the LC2 model in Section 3.5. Tests on the fitted time series processes from Figure 3.3a show that they are uncorrelated, which is a direct result of the identifiability constraint $\sum_t \kappa_t^{(1)} \kappa_t^{(2)} = 0$. However, we saw that the model period functions given in Figure 3.3b had a correlation of -75%, but gave exactly the same fitted mortality rates. We should therefore use multivariate processes

for both set of parameters.

Testing these parameters for stationarity, we find that both of the period functions in Figure 3.3 are non-stationary. We would therefore be justified in using a multivariate random walk for both sets of period functions (i.e., those from both Figure 3.3a and from Figure 3.3b).

We can see directly that this time series process is well-identified, since if

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t$$

then

$$\hat{\boldsymbol{\kappa}}_t = \hat{\boldsymbol{\kappa}}_{t-1} + A\boldsymbol{\mu} + A\boldsymbol{\epsilon}_t$$

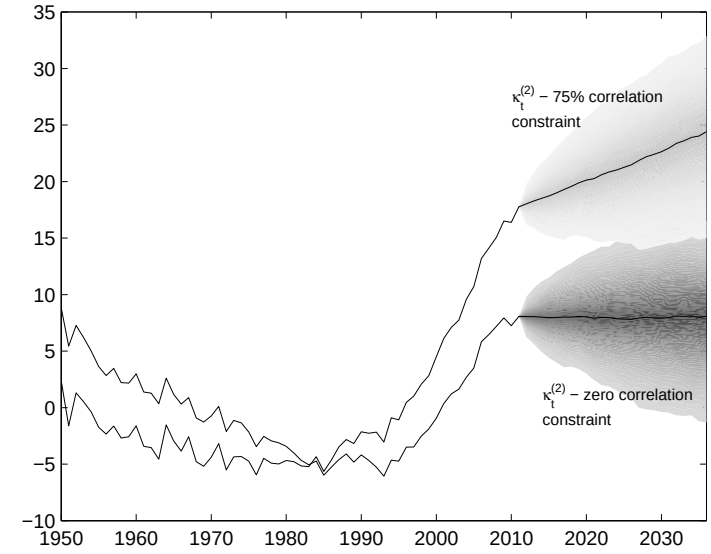
after applying the transformations in Equations 3.12 and 3.11. We see that integrated time series are unchanged by changes in the level of the period functions, and so are automatically invariant to the transformation in Equation 3.12.

At this point, it is also worth noting an important side effect of imposing orthogonality on the period functions in the LC2 model. $\kappa_t^{(1)}$ is usually found to be linear to quite a good approximation; so much so that this was called the *"universal pattern of mortality decline"* in Tuljapurkar et al. (2000). By construction, therefore, $\kappa_t^{(2)}$ cannot be roughly linear if we impose orthogonality, which makes projecting it trickier. We believe that this could be one of the reasons why the LC2 model is not more widely used, despite being a natural extension of the classic LC model. Often, the second term appears quadratic to quite a good approximation.[28] Various authors (such as Renshaw and Haberman (2003b) and Yang et al. (2010)) have suggested using break points or "hinges" in order to continue to use linear projection processes. However, this is a case of selecting a time series process specifically because of a feature of the period functions that is present solely because of the particular identifiability constraints imposed, and therefore the resulting projections will not be well-identified.

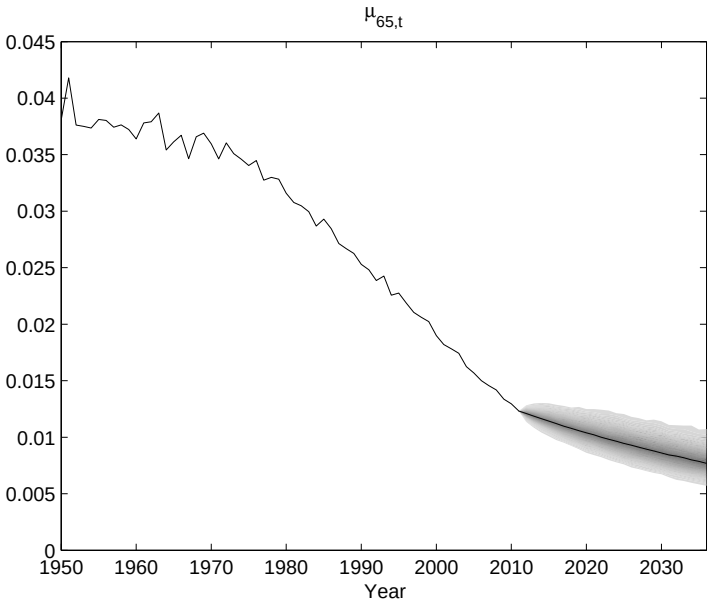
Using a multivariate random walk with drift for the time series processes in Figures 3.3a and 3.3b gives the projected $\kappa_t^{(2)}$ period functions in Figure 3.5a.[29] While these

[28] For instance in Renshaw and Haberman (2003b), Hatzopoulos and Haberman (2009) and Yang et al. (2010) as well as in Figure 3.3a.

[29] As seen in Figure 3.3, the difference between the two $\kappa_t^{(1)}$ parameters is very small.



(A) Projected LC2 $\kappa_t^{(2)}$



(B) Projected LC2 $\mu_{65,t}$

FIGURE 3.5: Projections from the LC2 model

projections appear quite different, the projected mortality rates from them at age 65, shown in Figure 3.5b are identical, thereby demonstrating that we have, indeed, chosen a well-identified projection method for the LC2 model.

3.9.3 Models with parametric age functions

In Section 3.6, it was shown that models with parametric age functions have subtly different identifiability issues when fitting them to data to those with non-parametric age functions. This is due to the transformations in Equation 3.11 not being allowed, since they changed the definition of the age functions and hence gave a different, but equivalent, model. However, we saw that this meant we could select between equivalent models, which had different definitions of the age functions, but gave identical fitted mortality rates. This was done in order to choose models with desirable properties such as a common normalisation scheme and orthogonal age functions. These subtle differences are also present when projecting the model.

First, the transformations in Equation 3.12 are used to impose a level on the period functions through identifiability constraints in models with parametric age functions in exactly the same manner as for models with non-parametric age functions. Consequently, we need to ensure that the time series processes used to project the period functions are identifiable under changes in location in exactly the same way as described for non-parametric age/period terms above. This means either using integrated time series processes or allowing for mean reversion to a non-zero level.

However, the transformations in Equation 3.11 are not needed in models with parametric age functions, since applying them would fundamentally change the model. Since we cannot normalise the age functions during the fitting process, we must instead define normalised (or self-normalising) age functions in advance. We cannot impose orthogonality on the age functions, although we could define orthogonal age functions a priori.

In addition, we cannot impose orthogonality on the period functions, as was done for the LC2 model, and therefore the period functions in models with parametric age functions will be correlated in general. This means that it is natural to project the period functions in such models using multivariate time series processes, just as we should in models with non-parametric age functions. However, because the transformations in Equation 3.11 are not applicable in models with parametric age functions, if we use a VARIMA(p,d,q) time series process for the period functions, as in Equation 3.28, we only have to ensure that the time series process is invariant to the transformation in Equation 3.12. To do

this, we substitute the transformed parameters, $\hat{\boldsymbol{\kappa}}_t = \boldsymbol{\kappa}_y + B$, into Equation 3.28 to find

$$\begin{aligned}\Delta^d \hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\mu} + \Delta^d B - \sum_{s=1}^{p} \Phi_s \Delta^d B + \sum_{s=1}^{p} \Phi_s \Delta^d \hat{\boldsymbol{\kappa}}_{t-s} + \sum_{r=0}^{q} \Psi_r \boldsymbol{\epsilon}_{t-r} \\ &= \hat{\boldsymbol{\mu}} + \sum_{s=1}^{p} \Phi_s \Delta^d \hat{\boldsymbol{\kappa}}_{t-s} + \sum_{r=0}^{q} \Psi_r \boldsymbol{\epsilon}_{t-r}\end{aligned}$$

Although the drift term, $\boldsymbol{\mu}$ has changed as a result of this transformation, the matrices Φ_s and Ψ_r have not. Consequently, we see that any structure we impose a priori upon the moving average and autocorrelation of the time series process is also unchanged by changes in the identifiability constraints in models with parametric age functions. This means that, in theory, it is possible to give each term distinct structure, such as different orders of integration or numbers of lags. This may be felt to be desirable if doing so gives projections with greater demographic significance.

For example, consider the exponential model in Equation 3.19. In this, we interpret $\kappa_t^{(2)}$ as representing the component of mortality change specific to very young ages, in excess of the changes in general mortality rates governed by $\kappa_t^{(1)}$. If we had a strong prior belief that these should mean-revert to a natural level (for instance, because we believed that infants should not receive systematically better or worse medical care than the general population), we might chose to allow our subjective demographic significance for the term to overrule a purely statistical evaluation of the time series process in this case. Because we do not use the transformation in Equation 3.11 to enforce a constraint when fitting the model, we do not have to ensure that our projection process is robust to its application when the model is projected.

We may also feel that such a restriction will give projected mortality rates with greater biological reasonableness. For example, we may have biological reasons for believing that infant mortality rates should always be higher than those for young children at age five, say. However, using a non-stationary time series process for $\kappa_t^{(2)}$ allows there to be scenarios with non-zero probability where this is violated, and therefore we might wish to use a stationary time series process for $\kappa_t^{(2)}$ to avoid any scenarios felt to be biologically unreasonable.[30]

However, such arguments ignore the fact that, for any model with parametric age functions, there are a range of equivalent models which give identical fitted mortality rates and so, ideally, should be projected using the same time series processes to give identical

[30] Similar arguments were considered in Cairns et al. (2006a) and Plat (2009a).

projected mortality rates.[31] There may also be features, such as changes in trend, which are present in the period functions for one model but absent in an equivalent model, and so are not objective features of the data. Since these equivalent models are linked by the transformation in Equation 3.11, it is still highly desirable to use general VARIMA processes, with no a priori structure placed on them, just as for models with non-parametric age functions.

In practice, it is not often that the demographic significance of a term in an AP mortality model leads to specific requirements about how it should be projected. For instance, while we may seek to rule out any possibility of mortality rates being lower at birth than at age five in the exponential model, this is highly unlikely to occur even if non-stationary time series processes are used for $\kappa_t^{(2)}$, since it is inconsistent with the historical data. We therefore recommend that general, well-identified, multivariate VARIMA processes are used to project the period functions in models with parametric age functions, unless these are shown experimentally to give biologically implausible projected mortality rates.[32]

3.9.4 Summary

In summary, we can say that in order to obtain projections which are well-identified from an AP model, we need to work backwards from our desire for time series processes which do not change form under the invariant transformations in Equations 3.11 and 3.12. This means that we should always use multivariate time series processes, as these support a unified approach to projection and allow us to check identifiability easily.

Identifiability also means, in general, that we should not treat the different period functions differently. In practice, this means assuming as little structure a priori for the time series processes as possible and using the same order of integration for each period function. In models with parametric age functions, however, there may be conflicts between

[31] As these are distinct models, this is a weaker requirement than is necessary to be well-identified under our definition above.

[32] In some circumstances, there are clear conflicts between the need for biological reasonableness in projected mortality rates and the desire to use the same time series processes for all period functions and in all equivalent models. These circumstances do not often arise in AP mortality models, but are more common in models with a cohort term which generates additional identifiability issues, and examples of such cases are discussed in Chapters 4 and 6. In such circumstances, it is usually preferable to choose processes which give biologically reasonable projections rather than identifiability under transformations which are not relevant in fitting the model.

achieving this and the biological reasonableness of the projected mortality rates. Treating the different period functions in the same manner is still highly desirable, however, as it avoids using different processes to project equivalent models, and often emerges naturally out of a statistical analysis of the fitted period functions. These conclusions are summarised in Table 3.1 below.

Property of time series process used in projection	Non-parametric age functions	Parametric age functions
Multivariate	Essential	Essential
Invariant to changes in scale	Automatic	Automatic
Invariant to changes in level (i.e., integrated or no preset level of mean reversion)	Essential	Essential
Correlation between period functions	Essential	Highly desirable
Have same order of integration	Essential	Highly desirable
Includes cross lags between period functions (if autoregressive)	Essential	Highly desirable

TABLE 3.1: Requirements for identifiable projection methods in AP mortality models

3.10 Conclusions

Most AP mortality models are not fully identified, since different sets of parameters will give identical fits to the observable data. This lack of identifiability requires us to impose additional constraints upon the parameters, which may help us interpret them and give them demographic significance. However, these additional constraints are chosen by the model user and therefore are subjective and arbitrary.

When using mortality models, it is important to be aware of all of the identification issues present and also how they need to be resolved. In many cases, this is done explicitly, such as in the model of Lee and Carter (1992). In others, it is done implicitly through the use of particular fitting procedures (e.g., Renshaw and Haberman (2003b) or Yang et al. (2010)). In cases where it is done implicitly, the identifiability constraints should still be clearly stated. This ensures that users of the model can correctly identify features of the fitted parameters which relate to the data (and so are worthy of investigation) and those which are merely artefacts of the identification scheme (such as the independence of the period functions in the LC2 model) and so are not. It also allows goodness of fit tests which use penalties based on the number of degrees of freedom in a model to be used reliably.

In addition, in parametric models, it is often desirable to select the age functions so that they have a consistent normalisation scheme based on a true norm, as this will allow comparisons to be made between the different age/period terms and will aid in the robustness of the projections. For models where the age functions have free parameters that are set with reference to the data, it is desirable to use self-normalising age functions to improve the stability of the numerical algorithms used to estimate the parameters and, hence, the model's robustness. However, these are properties of the age functions which are selected in advance of fitting the model, rather than being imposed during the fitting process via identifiability constraints.

These identification issues also have consequences when projecting the models. In general, in order to obtain identifiable projections, we should choose to project the model using multivariate processes which do not treat the period functions differently. It is also advisable to leave any vector representation of the time series as unstructured as possible (i.e., using general time series parameter matrices rather than imposing any structure on them a priori) in order for the representation to be robust across all identification schemes. Structure imposed through the arbitrary identifiability constraints will emerge when estimating these parameters. In models with parametric age functions, however, the use of identifiable projection methods is often desirable and natural, but may be subordinated to our desire for biological reasonableness in the projections.

In short, identification in AP mortality models is a non-trivial exercise which requires careful consideration and has consequences when we use the models to compare datasets or project future mortality rates. A lack of understanding of this can lead to projections which depend upon the arbitrary decisions made by the model user rather than the data. By understanding these issues, we can build more complex mortality models, for instance, via the "general procedure" of Chapter 5, and be confident that they are founded on a secure knowledge of the underlying mathematical structure of AP models. The subject of identifiability becomes considerably more complicated when we move beyond the AP structure to include the effects of year of birth (or cohort) as discussed in Chapter 4.

3.A Models without a static age function

As discussed in Chapter 2, a number of AP mortality models have been proposed which do not have an explicit static age function, α_x. These include the CBD model of Cairns et al. (2006a) and the model of Aro and Pennanen (2011), along with extensions of these. In order to achieve this, the age functions in the model must be parametric and therefore

known in advance of fitting the model to data. The structure of the AP model in this case is therefore

$$H = \beta\kappa$$

where $H = \{\eta_{x,t}\}$ as in Section 3.2.

In this case, we see that the identifiability issues in the model are simplified relative to the full structure in Equation 3.3. In particular, we see that the transformation in Equation 3.12 is no longer relevant and so the location of the period functions is no longer unidentified. Instead, the locations of the period functions are determined by the data and we no longer need to set them through identifiability constraints. Further, in the case where the age functions in the model are parametric, the transformation in Equation 3.11 is also no longer applicable, meaning that the model is fully identified. This is why no additional constraints are required for the models in Cairns et al. (2006a) and Aro and Pennanen (2011).

When projecting these models, we do not need to ensure that the time series processes are invariant to changes in the locations of the period parameters. However, since the fitted period parameters will have levels set by the data and these will typically be significantly different from zero, we need to allow for this possibility in our choice of time series processes. Consequently, in practice, time series processes which are either integrated or have the level of the period functions as a free parameter are often used to project the period functions. For instance, Cairns et al. (2006a) and Aro and Pennanen (2011) both used multivariate random walks with drift, which are invariant to changes in level even though this property is not strictly required.

Alternatively, some studies implicitly dispense with a static age function by fixing it in advance. For instance, Renshaw and Haberman (2003b) imposed

$$\alpha_x = \frac{1}{T}\sum_t \ln\left(\frac{d_{x,t}}{E^c_{x,t}}\right) \tag{3.29}$$

before estimating the other terms in the model. This sets the static age function as the average of observed mortality rates in the period. The value of the static age function is not subsequently revised when estimating the model.

In this case, the structure of the model becomes

$$\tilde{H} = \beta\kappa$$

where $\tilde{H} = \left\{\eta_{x,t} - \frac{1}{T}\sum_{\tau} \ln\left(\frac{d_{x,\tau}}{E^c_{x,\tau}}\right)\right\}$.

This means that Equation 3.12 is not an invariant transformation of the model and, consequently, the locations of the period functions are identifiable (i.e., defined by the data). Consequently, we do not need to then impose a constraint on the level of the period functions and, indeed, cannot do so without affecting the fitted mortality rates.

This is important when it comes to assessing the number of degrees of freedom in the mortality model, for instance, for the purposes of comparing the goodness of fit. For models where the level of the period functions is set via identifiability constraints, the model has $X + N(X + T)$ parameters and impose N level constraints and N^2 scale and orthogonality constraints on the model. In contrast, for models with a fixed static age function, the model has $N(X + T)$ free parameters and requires only the N^2 scale and orthogonality constraints. Therefore, models with a fixed static age function have $X - N$ fewer free parameters than might otherwise be expected. This was not allowed for in Haberman and Renshaw (2011) when comparing the goodness of fit for different models, which brings some of the conclusions of that study into question.

We also note that, in common with most statistical models with a two-stage estimation process (as discussed in Murphy and Topel (2002)), parameters estimated at the second stage may be biased and have distorted asymptotic distributions, compared with those estimated by a one-stage process. This is because of the hierarchical structure of the model: the second-stage parameters are only estimated conditional on the estimates of the first-stage parameters previously obtained, which are not known with certainty. To avoid this, we must either use a one-stage estimation process or use a bootstrapping procedure, such as those proposed in Brouhns et al. (2005) or Koissi et al. (2006) discussed in Section 3.8.1. These will allow fully for the uncertainty in both the parameters estimated at the first and second stages.

One reason for imposing the particular form of the static age function in Equation 3.29 is to give it approximately the same demographic significance as that which comes from using the constraint $\sum_t \kappa_t^{(i)} = 0$, i.e., that the static age function should represent the average mortality rate at each age over the period of the data, as shown in Equation

3.10. We might, therefore, expect to find

$$\sum_t \kappa_t^{(i)} \neq 0$$

for such a model. The difference between imposing the form of the static age function in Equation 3.29 and the estimate of the static age function found by maximising the fit to data and applying the identifiability constraint will depend on whether there are any systematic differences across periods between the fitted and observed mortality rates. We might, therefore, expect the difference between the two to be small if the model is a good fit to the data. Hence, for a model where the static age function is imposed, how different the value of $\sum_t \kappa_t^{(i)}$ is from zero is a measure of whether there are systematic differences between the observed and fitted mortality rates (i.e., whether there is structure remaining in the residuals from the model).[33] For models which do not provide an adequate fit to the data, there are likely to be systematic differences between the fitted and observed mortality rates and, hence, we will observe a value of $\sum_t \kappa_t^{(i)}$ further from zero if the static age function is imposed.

Nevertheless, even for a well-fitting model, it should be borne in mind that the period functions do possess an identifiable level when projecting them, even if this is small. It is therefore recommended that a non-zero level is allowed for in the time series processes used to project the period functions. In particular, we should not assume that any of the period functions mean-revert around zero, but, instead, allow them to mean-revert around an unspecified level. Nevertheless, this level would probably be close to zero, if the model is a good fit to the data, and could be tested for statistical significance (since it does not depend on an identifiability constraint).

In summary, models which either impose the value of the static age function a priori or which do not include an explicit static age function, have a reduced set of identifiability constraints compared with otherwise similar AP models where the static age function is unrestricted. Such models have levels for the period functions which are set with reference to the data rather than via an identifiability constraint. It is therefore necessary to include the period function levels when making projections from these models, even if the levels that have been estimated are close to zero. In most circumstances, they should therefore be treated in the same fashion as models with an explicit static age function. In contrast, models with no explicit age function but with a cohort term

[33] Indeed, if least squares methods are used to fit the model, the two are identical since this fitting procedure assumes that the residuals are independent and identically distributed.

possess different identifiability issues to comparable models with an explicit static age function, as discussed in Chapter 4.

3.B Maximal invariants

An alternative approach to using an arbitrary identification scheme was suggested by Nielsen and Nielsen (2014). This is to change the parameterisation of the model to an equivalent form with reduced dimensionality which does not suffer from identifiability issues. We can think of this reparameterisation as mapping the old parameters to a new set

$$g(\alpha, \beta, \kappa) = \{\tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}\}$$

The new parameters are chosen so that the new parameter space has the same dimension as the model space, $\mathcal{M}$, and so the mapping

$$\tilde{\Theta}(\tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}) = \Theta(g(\alpha, \beta, \kappa))$$

is injective (and so will not suffer from identification issues). The new parameters, $\{\tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}\}$, are known as "maximal invariant" parameters, since they are the set with the largest number of parameters (i.e., are "maximal"), and are injective and give the same fitted mortality rates as the original model in Equation 3.1 (i.e., the reparameterisation is "invariant").

As all of the maximally invariant parameters are freely varying (i.e., unconstrained) and $dim(\{\tilde{\alpha}, \tilde{\beta}, \tilde{\kappa}\}) = dim(\mathcal{M}) = X + N(X + T) - N(N + 1)$, we see that there are $X + N(X + T) - N(N + 1)$ parameters in the maximally invariant parameterisation. We can think of this as finding a parameterisation of the model which gives the same fit to data, but where every possible degree of freedom in the model is fully utilised in fitting the data.

Nielsen and Nielsen (2014) showed that one way that maximal invariant parameters can be used in the LC model in order to remove the lack of identifiability under the transformation in Equation 3.9 is through the use of the orthogonal complement to $\mathtt{1}$ (the $T \times 1$ column vector of ones defined in Section 3.2). This is a $T \times (T - 1)$ matrix, $\mathtt{1}_\perp$, used in Section 3.4, where every column is orthogonal to $\mathtt{1}$, i.e., $\mathtt{1}^\top \mathtt{1}_\perp = 0$.

Using the identity $\mathtt{I} = \mathtt{1}(\mathtt{1}^\top\mathtt{1})^{-1}\mathtt{1}^\top + \mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1}\mathtt{1}_\perp^\top$, we can decompose Equation 3.3 as

$$\begin{aligned} H &= \alpha\mathtt{1}^\top + \beta\kappa(\mathtt{1}(\mathtt{1}^\top\mathtt{1})^{-1}\mathtt{1}^\top + \mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1}\mathtt{1}_\perp^\top) \\ &= (\alpha + \beta\kappa\mathtt{1}(\mathtt{1}^\top\mathtt{1})^{-1})\mathtt{1}^\top + \beta(\kappa\mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1})\mathtt{1}_\perp^\top \\ &= \tilde{\alpha}\mathtt{1}^\top + \beta\tilde{\kappa}\mathtt{1}_\perp^\top \end{aligned} \tag{3.30}$$

where $\tilde{\kappa}$ is now a $N \times (T-1)$ matrix. We can see that if we transform the original parameters using Equation 3.12 we obtain

$$\begin{aligned} \tilde{\hat{\kappa}} &= \hat{\kappa}\mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1} \\ &= (\kappa + B\mathtt{1}^\top)\mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1} \\ &= \kappa\mathtt{1}_\perp(\mathtt{1}_\perp^\top\mathtt{1}_\perp)^{-1} \\ &= \tilde{\kappa} \end{aligned}$$

i.e., the lack of injectivity in the model is now between the mapping from the old parameterisation to the new, but the transformation of the new parameters to the fitted mortality rates is injective. This has explicitly reduced the number of parameters in the model from $X+N(X+T)$ to $X+N(X+T-1)$ and means that the revised $\tilde{\kappa}$ parameters have identifiable location. However, the parameters are still not fully identified under the transformations in Equation 3.11, and therefore the maximally invariant reparameterisation has not completely solved the identifiability issues in the model.

It is also apparent that this technique does not depend on the form of the matrix β. Specifically, if we use parametric age functions, then we can still use the same analysis to remove the lack of identifiability in the level of the period functions.

Mathematically, the approach suggested in Nielsen and Nielsen (2014) is very elegant. However, in practice, the approach has hidden rather than removed the lack of identifiability to the transformations in Equation 3.12. This is because $\mathtt{1}_\perp$ is not unique, but can be chosen by the model user. The model user's choice does not have any statistical consequences and is equivalent to choosing a basis in the $(T-1)$ dimensional orthogonal subspace of $\mathbb{R}^T$ spanned by $\mathtt{1}_\perp$. Nonetheless , this choice will have consequences when we come to interpret the demographic significance and project the parameters in the model.

For instance, we might choose

$$\mathtt{1}_\perp = \begin{pmatrix} -1 & 0 & 0 & \cdots \\ 1 & -1 & 0 & \\ 0 & 1 & -1 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{pmatrix} \tag{3.31}$$

This choice means that $(\kappa \mathtt{1}_\perp)_t^{(i)}$ corresponds to $\Delta\kappa_t^{(i)} = \kappa_t^{(i)} - \kappa_{t-1}^{(i)}$, the first differences between successive period parameters, which is invariant to change in the level of $\kappa_t^{(i)}$. This has a natural interpretation and is related to modelling "mortality improvement rates" as was done in Haberman and Renshaw (2012) and Mitchell et al. (2013). Alternatively, we could choose

$$\mathtt{1}_\perp = \begin{pmatrix} -1 & -1 & -1 & \cdots \\ 1 & 0 & 0 & \\ 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ \vdots & & & \ddots \end{pmatrix} \tag{3.32}$$

This choice implies that $(\kappa \mathtt{1}_\perp)_t^{(i)}$ corresponds to $\kappa_t^{(i)} - \kappa_1^{(i)}$, the changes in the period function from its initial value. This is also invariant to change in the level of $\kappa_t^{(i)}$, but will have a very different pattern from that of the first differences used previously (and be projected using different methods). We could consider these choices as analogous to the imposition of the identifiability constraints $\sum_t \kappa_t^{(i)} = 0$ and $\kappa_1^{(i)} = 0$, respectively. Most statistical packages will select a $\mathtt{1}_\perp$ matrix using a numerical algorithm and so $\kappa \mathtt{1}_\perp$ will not have a natural interpretation, limiting the demographic significance of any maximally invariant parameters.

When we come to project the model, we will need to extend $\mathtt{1}_\perp$ as well as $\tilde{\kappa}_t$. For instance, to project τ years into the future, we will need to generate a $((T+\tau)\times(T+\tau-1))$ matrix $\tilde{\mathtt{1}}_\perp$. However, in order to be consistent with the fitted mortality rates, we will also need to ensure that the $(T \times (T-1))$ upper left submatrix of $\tilde{\mathtt{1}}_\perp$ is identical to the matrix $\mathtt{1}_\perp$ used when fitting the model. This may not be the case when using some common algorithms to generate these orthogonal matrices, leading to inconsistencies between the fitted and projected mortality rates, and so it is important that we understand the method used to generate orthogonal matrices in order to ensure consistency.

Even more problematic, our choice of $\mathtt{1}_\perp$ might not preserve the time ordering of κ_t. For instance, we can re-order the columns of the $\mathtt{1}_\perp$ matrix in Equation 3.31, so that $(\kappa \mathtt{1}_\perp)^{(i)}$ is still a row vector of the first differences in $\kappa_t^{(i)}$ but not in chronological order. Since it is the time-ordering of $\kappa_t^{(i)}$ which allows us to interpret it as a time series and project it into the future in order to forecast mortality rates, this is highly undesirable.

Furthermore, we have not removed the lack of identifiability under the transformations in Equation 3.11. We therefore will still need to impose a normalisation scheme on the age/period terms and can select orthogonal age functions using this transformation. Hence, much of the discussion in Section 3.9 is still relevant, even using a choice for $\mathtt{1}_\perp$ which preserves the time ordering of κ_t.

In summary, the use of maximal invariants in AP mortality models has a number of elegant mathematical properties. However, moving to this framework involves losing much of the demographic significance associated with the parameters in a standard AP mortality model and does not solve many of the key issues with projecting such models. It is, therefore, unlikely that such an approach will be suitable for the purposes of most users of mortality models.

Chapter 4

Identifiability in Age/Period/Cohort Mortality Models

4.1 Introduction

Many modern models of mortality include parameters to capture the impact of lifelong mortality effects which follow individuals from birth, building on the findings of studies such as Wilmoth (1990) and Willets (1999, 2004). Understanding such "cohort" effects can be of critical importance, especially for those interested in understanding the mortality experience of a specified group of lives, such as members of a pension scheme or policyholders in an annuity book. Examples of models incorporating cohort parameters include those proposed in Renshaw and Haberman (2006), Cairns et al. (2009), Plat (2009a), O'Hare and Li (2012a), Börger et al. (2013) and Chapter 5.

In Chapter 2, we argued that the time has come to undertake a more holistic analysis of the class of age/period/cohort (APC) models and began this analysis by outlining their common structure. In Chapter 3, we focused on the subset of this class without a cohort term, namely on age/period (AP) models, and examined their identifiability issues.

We found that, for AP models, there are a number of "invariant transformations" which change the parameters, but not the fitted mortality rates. The existence of these transformations lead to identifiability issues, meaning that there are certain features of the parameters in a model which are not defined by the data. Instead, they are only determined by the arbitrary identifiability constraints we impose, and therefore have no independent meaning. Consequently, we must be careful to ensure that our results from

using mortality models do not depend upon these features of the parameters. These issues with identifiability can lead to models which lack robustness when fitted to data, cause us to draw faulty and erroneous conclusions when analysing the historical data, and bias our projected mortality rates in future. We also found that, unless we choose our projection methods carefully, our projections of mortality can depend upon the arbitrary choice of identifiability constraint. This should be avoided, so we discussed how to choose projection methods which give "well-identified" projections of mortality rates.

The addition of a set of cohort parameters to a mortality model can generate additional identifiability issues which are fundamentally unlike anything present in otherwise similar AP models. These are caused by the collinearity between age, period and cohort. In the context of the APC mortality models discussed in this study, we find that certain deterministic trends found within the fitted parameters are unidentifiable by the models, and therefore do not possess any meaning other than that imposed by our arbitrary identifiability constraints. This, in turn, means that it is both more important and more difficult to ensure that projections from these models are well-identified, as we must separate these unidentified trends (which depend entirely upon the identifiability constraints) from the variation around the trends, which is meaningful and needs to be projected consistently with what has been observed in the past. Thus, although the present study extends the work of Chapter 3, it is necessary to view the underlying identifiability issues in a fundamentally different way and, consequently, develop a new set of tools to solve them.

In this chapter, we study the identifiability issues present in some of the simplest APC models in order to demonstrate the problems in action and their potential resolution. In these simple cases, the identifiability issues can appear trivial, and their impact on our analysis of historical and projected mortality rates relatively minor. However, we believe that it is vital to fully understand these issues in the context of simple models, since they become considerably more important in more complicated models. Indeed, recognising these issues and solving them was vital to the development of the "general procedure" for constructing APC mortality models, described in Chapter 5, and appropriately projecting such models, as we discuss in Chapters 6, 7 and 8.

The outline of the chapter is as follows. Section 4.2 reviews the structure of general APC mortality models described in Chapter 2. Section 4.3 introduces the concept of identifiability in the context of the simplest and most widely used APC model and develops our understanding of how cohort effects create fundamentally new identification

issues in this model compared with the simpler AP model. Section 4.4 generalises this by examining the issue of identifiability in more general APC models with parametric age functions. Section 4.5 investigates the consequences of identification for projection, first by looking at the model discussed in Section 4.3 and then in a more general case. Finally, Section 4.6 concludes.

4.2 Structure of age/period/cohort models

An APC mortality model is one which assumes that mortality rates can be modelled as a series of terms involving functions of age, x, period, t, and year of birth, $y = t - x$.[1] This can be written as

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x} \tag{4.1}$$

where

- $\eta_{x,t}$ is a link function to transform the response variable into a form suitable for modelling and linking it to the proposed predictor structure;
- α_x is a static function of age;[2]
- $\kappa_t^{(i)}$ are period functions governing the evolution of mortality with time;
- $\beta_x^{(i)}$ are age functions modulating the impact of the period function dynamics over the age range; and
- γ_y is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as as it ages.

We also note that the general APC mortality model in Equation 4.1 can be re-written as

$$\eta_{x,t} = \alpha_x + \boldsymbol{\beta}_x^{\top} \boldsymbol{\kappa}_t + \gamma_{t-x} \tag{4.2}$$

where

$$\boldsymbol{\kappa}_t = \left(\kappa_t^{(1)}, \quad \dots \quad \kappa_t^{(N)}\right)^{\top}$$
$$\boldsymbol{\beta}_x = \left(\beta_x^{(1)}, \quad \dots \quad \beta_x^{(N)}\right)^{\top}$$

[1] In this study, we assume that $x \in [1, X]$ and $t \in [1, T]$ and hence that years of birth, y, are in the range $(1 - X)$ to $(T - 1)$. In practice, x and t will be given by the range of the data being used.

[2] We consider models of the form of Equation 4.1 but without a static age function in Appendix 4.B.

This form is useful when projecting these models, as discussed in Section 4.5.

The general structure of APC models was discussed in detail in Chapter 2. In particular, we found that APC mortality models have different demographic significance[3] depending on whether the age functions $\beta_x^{(i)}$ are non-parametric[4] or parametric.[5]

In Chapter 3, we used linear algebra to analyse the structure of AP mortality models as mappings from a space of parameters to a model space, and found that in order for these mapping to be unique, the spaces had to have the same dimension. In addition, AP models can be sub-divided into those with parametric age functions and those where the age functions are non-parametric. While the two families have similar identifiability issues, these needed to be solved using different methods in order to preserve the demographic significance of the parametric age functions.[6] It is important to note that AP mortality models are nested within the class of APC models, and, therefore, all of the issues raised in Chapter 3 are still applicable for APC mortality models.

APC models have additional identifiability issues which are fundamentally different from anything present in otherwise similar AP models, hence alternative methods are necessary to analyse them. They are caused by the collinearity between the dimensions of age, period and cohort, because period = year of birth + age. This gives us the freedom to re-write functions of cohort as functions of age and period, or vice versa. The additional identifiability issues generated by the cohort term depend fundamentally on the definition of the age functions within the model, and so are specific to the model in question. We find that APC models with non-parametric age functions do not have any extra identifiability issues beyond those discussed for AP models in Chapter 3, as shown in Appendix 4.A. Models with certain types of parametric age functions require additional identification as discussed in Section 4.4.

In Chapter 2, we also found that difficulties with estimating and assigning demographic significance to the cohort parameters mean that, in practice, most models use only one

[3] Demographic significance is defined in Chapter 2 as the interpretation of the components of the model being explainable in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates.

[4] The values of the age functions $\beta_x^{(i)}$ at different ages x are fitted without any a priori structure or functional form. See Chapter 2.

[5] The age functions $\beta_x^{(i)}$ take a specific functional form $\beta_x^{(i)} = f^{(i)}(x;\theta^{(i)})$, defined in advance of fitting the model to data. For simplicity, the dependence of the age functions on $\theta^{(i)}$ is suppressed in the remainder of this chapter.

[6] These different methods are not germane to the arguments in this study. Interested readers should consult Chapter 3.

cohort term (without any modulating age function) and do not involve any age/cohort interactions for reasons of both simplicity and robustness. We follow the same approach in this chapter, and so do not consider models such as that proposed in Renshaw and Haberman (2006) or Model M8 in Cairns et al. (2009).

4.3 Identifiability in the classic APC model

The simplest APC model (referred to here as the "classic APC model") has a long history and is widely used in the fields of medicine, epidemiology and sociology as well as in demography and actuarial science.[7] It has the following form

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \tag{4.3}$$

It can be seen that the classic APC model has one age/period term with $f(x) = 1$, which is parametric in the sense defined in Chapter 2.

A model is fully identified when all the parameters in it can be uniquely determined by reference to the available data. In contrast, the classic APC model (as with most APC models) is not fully identified, because there exist different sets of parameters which will give the same fitted mortality rates and consequently the same goodness of fit for any data set. This phenomenon is not unique to APC mortality models. However, it is very widespread in such models and has significant implications when we come to make projections using them.

The issue of identifiability in the classic APC model also has a very long history.[8] It is, therefore, a good starting point to determine whether the issues raised in identifying the parameters in Equation 4.3 can be generalised to the more complex APC models used in mortality modelling. We can see that this model is not fully identified, since if we use the transformations in Equations 4.4, 4.5 and 4.6 to obtain new sets of parameters, we

[7] For instance, see Hobcraft et al. (1982), Osmond (1985), O'Brien (2000), Carstensen (2007) and Kuang et al. (2008b).

[8] For instance, see Glenn (1976), Fienberg and Mason (1979), Rodgers (1982), Holford (1983), Clayton and Schifflers (1987), Wilmoth (1990), Yang et al. (2004), Kuang et al. (2008a) and O'Brien (2011).

do not change the fitted mortality rates and hence the fit to the data

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - a, \kappa_t + a, \gamma_y\} \tag{4.4}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - b, \kappa_t, \gamma_y + b\} \tag{4.5}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x + c(x - \bar{x}), \kappa_t - c(t - \bar{t}), \gamma_y + c(y - \bar{y})\} \tag{4.6}$$

where a bar denotes the arithmetic mean of the variable over the relevant data range.[9] We call such transformations "invariant" for this reason. The existence of invariant transformations means that the model possesses identifiability issues, because no one set of parameters is determined uniquely from the data.

The transformation in Equation 4.6 is fundamentally unlike any of the transformations present in AP models discussed in Chapter 3, since it involves functions of age, period and year of birth rather than constants. It is a consequence of the collinearity between these dimensions, $y = t - x$, which enables us to decompose a linear function of year of birth into linear functions of age and period, and vice versa. This transformation generalises for many, more complex APC models with parametric age/period terms, as we discuss in Section 4.4.

We say that linear trends in the data are "unidentifiable" by the model, that is, they cannot be uniquely apportioned to either age, period or year of birth (as was discussed in Wilmoth (1990)). The linear trends observed in the parameters of the classic APC model therefore have no independent meaning, as different sets of parameters, with different linear trends will give exactly the same observable quantities such as fitted mortality rates.

The existence of unidentifiable linear trends in the classic APC model is of practical as well as theoretical importance. This is because we often see features of the (transformed) mortality rates which are approximately linear in age and time. For instance, the shape of the age function, α_x, is approximately linear at high ages,[10] whilst κ_t is often approximately linear.[11] The structure of the model means that we are fundamentally unable to separate these linear trends from a linear trend in the cohort parameters.

[9] e.g., $\bar{x} = \frac{1}{X}\sum_x x = 0.5(X + 1)$.

[10] If $\eta_{x,t} = \ln(\mu_{x,t})$, this is the Gompertz model, whilst if $\eta_{x,t} = \text{logit}(q_{x,t})$, this is the Perks model for mortality.

[11] See, for instance, Tuljapurkar et al. (2000), who went so far as to call this the *"universal pattern of mortality decline"*.

Because different sets of parameters give the same fit to the data, we cannot use the data to apportion the linear trend to either the age, period or cohort terms. One method of solving this issue is to move to a "maximally invariant" set of parameters, as discussed in Kuang et al. (2008a) and Nielsen and Nielsen (2014), which involves reparameterising the model in an equivalent form with reduced dimensionality, which avoids the identifiability issues. This approach is discussed in Appendix 4.C.

An alternative and much more common approach is to impose additional identifiability constraints on the parameters in order to specify them uniquely.[12] These constraints manually apportion the linear trend between the different terms in the model. Imposing suitable constraints on the model involves the selection of a single set of parameters from the family of equivalent parameter sets, all of which give identical fitted mortality rates. In this sense, the manual apportionment is arbitrary - it does not depend upon any observable property of the data, but is a product of the model user's subjective interpretation of the demographic significance of the parameters.

For example, one set of identifiability constraints is $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = 0$.[13] These identifiability constraints allow us to impose our interpretation of the demographic significance of the parameters onto the model. For example, the first two of the constraints above mean that α_x can be interpreted as an "average" level of mortality at age x, over the period, with κ_t and γ_y representing deviations from this average level. The third constraint requires that there are no deterministic linear trends within the fitted cohort parameters, since any linear trend in these parameters will be arbitrarily assigned to the age and period effects by using the transformation in Equation 4.6. This is in line with the demographic significance we assign to the cohort parameters in Chapter 2.

However, it is important to note that these additional identifiability constraints are arbitrary. For instance, the constraints $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y (y - \bar{y}) = 0$ (used later in Section 4.5.2) could also be imposed and would give different estimated parameters with exactly the same fit to data and have the same demographic significance. Further, the choice of having no linear trend in the cohort parameters does not have any independent meaning, since it is entirely dependent upon the identifiability constraints chosen. While these constraints might allow us to interpret the demographic

[12] We say that the transformations in Equations 4.4, 4.5 and 4.6 cause issues with the *identifiability* of the model. *Identification* of the model is accomplished by imposing a set of identifiability constraints and using the invariant transformations to achieve these constraints.

[13] Here n_y is the number of observations of cohort y in the data and so $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x}$.

significance of the parameters, this interpretation nevertheless depends entirely on the user's judgement rather than on the underlying data. For instance, a different choice of identifiability constraints could be used to impose that the period parameters, κ_t, had no linear trend, which would give the parameters a different demographic significance but leave the fitted mortality rates unchanged. We must, therefore, take care to ensure that our projections of observable quantities such as mortality rates do not depend on our arbitrary identification scheme, as discussed in Section 4.5.

4.4 Identifiability in APC models with parametric age functions

Many of the more complex APC mortality models being proposed contain cohort parameters in the same form as in the classic APC model (i.e., without an age modulating $\beta_x^{(0)}$ function). Cairns et al. (2009) and Haberman and Renshaw (2011) found that models with a cohort term fit the data better than otherwise similar AP models, especially for the UK population, where a strong cohort effect has been observed by Willets (1999, 2004) and others. It is therefore natural to ask whether the additional issues with identifiability present in the classic APC model are also present in these more complex models.

In Appendix 4.A, we show that APC models with non-parametric age functions do not possess any additional, non-trivial identification issues, beyond those found in similar AP models discussed in Chapter 3. We have already seen, however, that in the simplest case of the classic APC model, the additional structure in the model caused by having a parametric age function combined with the collinearity of age, period and cohort can yield new identification issues.

For a general model with parametric age functions

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} + \gamma_{t-x} \tag{4.7}$$

we can try to generalise Equation 4.6 to look for invariant transformations of the form

$$\{\hat{\alpha}_x, \hat{f}^{(i)}(x), \hat{\kappa}_t^{(i)}, \hat{\gamma}_y\} = \{\alpha_x - a(x), f^{(i)}(x), \kappa_t^{(i)} - k^{(i)}(t), \gamma_y + g(y)\} \tag{4.8}$$

where $a(x)$, $k^{(i)}(t)$ and $g(y)$ are smooth functions.[14] Because the formulae used for the age functions define the model being used, in the sense of Chapter 2, we desire that they do not change under the invariant transformations, i.e., $\hat{f}^{(i)}(x) = f^{(i)}(x)$. Transformations which changed the age functions in the model would give a fundamentally different model, albeit one which gave the same fit to the data. In Chapter 3, we called different models, with different definitions of the age functions, that gave identical fits to the data "equivalent models".

In order for the transformation in Equation 4.8 to leave Equation 4.7 unchanged, we require

$$g(t-x) = a(x) + \sum_{i=1}^{N} f^{(i)}(x)k^{(i)}(t) \tag{4.9}$$

If this is true, we say that the deterministic trends $k^{(i)}(t)$ and $g(y)$ are "unidentifiable", since the model is unable to apportion them between the age/period and cohort terms, in the same way as with the unidentifiable linear trends in the classic APC model. Instead, we must manually apportion these trends by means of additional identifiability constraints. These deterministic trends in the fitted parameters, therefore, lack any objective meaning, since they are entirely dependent on the choice of identifiability constraints. Nevertheless, they must be allowed for when projecting the APC mortality model, as discussed in Section 4.5, even if they appear to be comparatively small.

The first thing to note from Equation 4.8 is the trivial case where Equation 4.9 holds, i.e., $g(y) = a(x) = b$, a constant, and $k^{(i)}(t) = 0$, $\forall t$. This is simply a transformation of the form in Equation 4.5. It does not involve any age/period terms and so holds for all APC models, including those with non-parametric age functions.

To find less trivial transformations, we take a Taylor expansion of $g(y)$ around $-x$, assuming that it is an infinitely differentiable function of year of birth

$$g(t-x) = g(-x) + \sum_{j=1}^{\infty} \frac{1}{j!} t^j \left.\frac{d^j g}{dy^j}\right|_{y=-x} \tag{4.10}$$

Comparing this to Equation 4.9, we can set $a(x) = g(-x)$ and $k^{(j)}(t) = \frac{1}{j!}t^j$ if $f^{(j)}(x) = \left.\frac{d^j g}{dy^j}\right|_{y=-x}$, i.e., the derivatives of g are a subset of the age functions of the model. Models

[14] While, α_x and κ_t are only defined for integer x and t, the parametric age functions $f^{(i)}(x)$ are defined for continuous x and so it make sense to look for transformations which also use continuous functions, as in the classic APC model in Section 4.3.

of the form in Equation 4.7 have a finite number, N, of age/period terms and, therefore, we require that $g(y)$ has a finite series of derivatives. There are two cases when g will have a finite sequence of derivatives, either

1. the derivatives terminate after $M \leq N$ terms say, or
2. the form of the derivatives is cyclical so that $\frac{d^{j+M}g}{dy^{j+M}}\Big|_{y=-x} = K \frac{d^j g}{dy^j}\Big|_{y=-x}$ for some integer $M \leq N$ and constant K.

4.4.1 Polynomial age functions

For the Taylor series to terminate in a finite number of terms, we require that $\frac{d^j g}{gy^j} = 0$, $\forall j > M$, and therefore that $g(y)$ must be a polynomial in y of order M.

Theorem 4.1. *APC mortality models of the form in Equation 4.1 and age functions spanning the polynomials to order $M-1$ possess invariant transformations which add a polynomial of order M to the cohort function.*

Sketch of Proof Take $g(y)$, a general polynomial of order M, and expand as a function of x and t. This can then be regrouped into an equivalent form that corresponds to the age/period terms in the model, in order to see how $g(y)$ can be absorbed into the age/period structure

$$
\begin{aligned}
g(y) &= \sum_{n=0}^{M} a_n y^n \\
\Rightarrow g(t-x) &= \sum_{n=0}^{M} a_n (t-x)^n \\
&= \sum_{n=0}^{M} a_n \sum_{m=0}^{n} \binom{n}{m} t^m (-x)^{n-m} \\
&= \sum_{n=0}^{M} a_n \left[(-x)^n + \sum_{m=1}^{n} \binom{n}{m} t^m (-x)^{n-m} \right] \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{n=1}^{M} \sum_{l=0}^{n-1} a_n \binom{n}{l} t^{n-l} (-x)^l \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{l=0}^{M-1} (-x)^l \sum_{n=l+1}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{n=0}^{M} a_n (-x)^n + \sum_{l=0}^{M-1} (-1)^l f^{(l)}(x) \sum_{n=l+1}^{M} a_n \binom{n}{l} t^{n-l} \\
&= a(x) + \sum_{l=0}^{M-1} f^{(l)}(x) k^{(l)}(t)
\end{aligned}
$$

If there are age functions in the model of the form $f^{(j)}(x) = x^j$ of $j = 0, 1, \ldots M-1$, the expression above corresponds to Equation 4.9 with $a(x) = \sum_{n=0}^{M} a_n(-x)^n$ and $k^{(j)}(t) = (-1)^j \sum_{n=j+1}^{M} a_n \binom{n}{j} t^{n-j}$. More generally, we only require that the age functions span the first $M-1$ polynomials, because these are equivalent to a model with $f^{(j)}(x) = x^j$ such as that in the derivation above. □

We can think of the transformation as expanding the polynomial $g(y)$ into terms in x and t, grouping these and then combining them with the appropriate age/period terms. A model with age functions spanning the first $M-1$ polynomials therefore has an additional $M+1$ degrees of freedom (represented by the coefficients, a_n, of the general polynomial) which do not affect the fit to the data. This is similar to the analysis in Wilmoth (1990), which argues that higher order polynomial trends in the cohort parameters will cause identifiability problems in a mortality model if sufficient age/period terms of suitable form exist within the model. These additional degrees of freedom mean that we need to impose an additional $M+1$ identifiability constraints, which assign the $M+1$ unidentifiable polynomial trends between the different age/period and cohort terms in the model.

The simplest example of this is the transformation of the classic APC model described in Section 4.3. This has a single parametric age function $f(x) = 1$ which spans the polynomials to order 0. The model will then allow first order polynomials (i.e., linear terms) to be added to the cohort parameters with offsets made to the static life function and the period term without changing the fitted mortality rates. These are exactly the invariant transformations described in Equations 4.5 and 4.6. Consequently, we impose two additional identifiability constraints for the cohort parameters in the model to identify their level and linear trend.

4.4.1.1 The Plat models

In Plat (2009a), two new APC mortality models were introduced. These can be written[15]

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_{t-x} \tag{4.11}$$

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{4.12}$$

The second of these models was introduced as a simplification of the first, with the expectation that it would be more suitable for modelling mortality at high ages. We call the model in Equation 4.11 the "Plat model" and the model in Equation 4.12 the "reduced

[15] We define $x^+ \equiv max(x, 0)$.

Plat model" for this reason.[16]

The first point to note is that both the Plat and reduced Plat models nest the classic APC model, and therefore the invariant transformations in Equations 4.4, 4.5 and 4.6 are also applicable for both models.

The second point to note is that these models also nest simple AP mortality models,[17] and therefore the results of Chapter 3 are still applicable. This means that the "locations" of the period functions are undefined and need to be identified by imposing a constraint on their levels. Usually this is of the form

$$\sum_t \kappa_t^{(i)} = 0$$

These invariant transformations were noted by Plat (2009a) and used to impose suitable identifiability constraints.

However, the third point to note is that both of these models have age functions $f^{(1)}(x) = 1$ and $f^{(2)}(x) = (x - \bar{x})$ which span the polynomials to linear order. Using the result of Theorem 4.1, we should be able to find a transformation of the parameters which adds a quadratic polynomial in y to the cohort parameters, but leaves the fitted mortality rates unchanged. Indeed, we find that the transformation

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{&\alpha_x - d(x-\bar{x})^2, \\ &\kappa_t^{(1)} - d(t-\bar{t})^2, \kappa_t^{(2)} + 2d(t-\bar{t}), \gamma_y + d(y-\bar{y})^2\}\end{aligned} \tag{4.13}$$

leaves the fitted mortality rates unchanged for both the Plat and reduced Plat models. We say that these models have unidentifiable quadratic trends, which have to be manually allocated between the different parameters via identifiability constraints.

Hence, we require three identifiability constraints on the cohort parameters in the Plat and reduced Plat models, i.e., to apportion the level, linear trend and quadratic trend between the different age/period and cohort terms, plus identifiability constraints on the levels of the period functions. This means that for full identification of the models, we require an additional identifiability constraint to those discussed in Plat (2009a).

[16] This model can also be thought of as an extension to model M6 in Cairns et al. (2009), with a static age function, or as an extension to the "CBDX" model discussed in Chapter 3 with a cohort term.

[17] In particular, both models nest the "CBDX" model discussed in Chapter 3.

If the model user fails to allocate the quadratic trend between the different terms via an additional identifiability constraint, then the fitting algorithm will make an apportionment in order to achieve convergence. However, this apportionment will not be based on any particular desired demographic significance and will depend on the specific details of fitting algorithm, such as the starting parameter values used. To illustrate, instead of removing quadratic trends from the cohort parameters and apportioning them to the age/period terms, the fitting algorithm may split any quadratic trends between the cohort parameters and the age/period terms, giving values of γ_y with an apparent quadratic trend in y. Not only is this contrary to our desired demographic significance, it can make comparing parameters across datasets difficult due to the presence or absence of quadratic trends which do not have any meaning independent of the data.

In addition, a failure to fully identify the model can lead to inefficient fitting algorithms, which take a long time to converge to a solution, as discussed in Hunt and Villegas (2015). Furthermore, they can also give parameter estimates which are not robust to small changes in the data (e.g., an additional year of data), since such changes can cause the fitting algorithm to abruptly change the allocation of the unidentifiable trends. For these reasons, it is very important to ensure that the APC mortality models we use are fully identified by imposing sufficient identifiability constraints to uniquely estimate all the parameters in the model.

Following the same approach as used for the classic APC model, we might choose to impose the constraints in Section 4.3 and extend these to impose $\sum_y n_y(y-\bar{y})^2\gamma_y = 0$ to remove quadratic trends in the cohort parameters and apportion them to the age/period terms. However, as with the classic APC model, this choice is arbitrary and a different choice of constraints will make no difference to the fitted mortality rates, only to the interpretation we give to the parameters.

In Section 4.3, we saw that the lack of identifiability of the linear trends in the model, due to the transformation in Equation 4.6, was of practical as well as theoretical importance because linear trends were often observed in both the age and period terms. Similarly, the transformation in Equation 4.13 is of practical importance when fitting the Plat model, because we usually see some curvature in α_x at high ages and also systematic departures from the linearity of the period functions.[18] These quadratic trends will, therefore, not be distinguishable from a quadratic trend in the cohort parameters

[18] For instance, see Booth et al. (2002), who curtailed the use of the data in the Lee and Carter (1992) model based on when a linear assumption for κ_t is no longer appropriate.

in the Plat model. However, because the observed magnitude of such trends is typically smaller than the linear trends observed in the age/period terms, failure to fully identify the quadratic trend in the data will typically have a lower, though still important, impact than a failure to identify the linear trend.

It is worth noting that the transformation in Equation 4.13 does not treat the different period functions equally, i.e., a term which is quadratic in t is added to $\kappa_t^{(1)}$, a term linear in t is added to $\kappa_t^{(2)}$, whilst $\kappa_t^{(3)}$ is unchanged by the invariant transformation for the Plat model. However, this is true only for the particular definition of the age functions shown. To illustrate, instead of the Plat model in Equation 4.11, we could instead have chosen an equivalent model of the form

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})^+ \kappa_t^{(2)} + (\bar{x} - x)^+ \kappa_t^{(3)} + \gamma_{t-x} \tag{4.14}$$

Such a model will trivially give the same fitted mortality rates as that in Equation 4.11 and has the same number of parameters, and so will have the same number of identifiability issues. However, the transformation corresponding to Equation 4.13 for this model will now add terms linear in t to both $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$. Specifically, for this model, we have the invariant transformation

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{&\alpha_x - d(x - \bar{x})^2, \kappa_t^{(1)} - d(t - \bar{t})^2, \\ &\kappa_t^{(2)} - 2d(t - \bar{t}), \kappa_t^{(3)} + 2d(t - \bar{t}), \gamma + d(y - \bar{y})^2\}\end{aligned} \tag{4.15}$$

in contrast to the transformation in Equation 4.13. Specifically, we note that whilst the transformation in Equation 4.13 did not involve $\kappa_t^{(3)}$, the transformation in Equation 4.15 does. The invariant transformations of the model are therefore specific to the age functions present, and may be different in different models, even if those models give an equivalent fit to data.

4.4.2 Exponential and trigonometric age functions

The other case where Equation 4.10 potentially yields invariant transformations of the parameters occurs when the derivatives of $g(y)$ are cyclical with period $M \leq N$.

Theorem 4.2. *APC mortality models of the form in Equation 4.1 with exponential or trigonometric age functions possess invariant transformations which add similar exponential or trigonometric functions to the cohort parameters.*

Sketch of Proof In order for the derivatives of $g(y)$ to be cyclical with period M, we require

$$\frac{d^M g}{dy^M} = Kg \tag{4.16}$$

for some non-zero constant K. Substituting this into Equation 4.10 and comparing with Equation 4.9 gives

$$\begin{aligned} g(t-x) &= \sum_{j=0}^{M-1} \left.\frac{d^j g}{dy^j}\right|_{y=-x} \sum_{k=1}^{\infty} \frac{1}{(j+kM)!} t^{j+kM} \\ &= \sum_{j=0}^{M-1} f^{(j)}(x)k(t) \end{aligned}$$

This is of the form of Equation 4.9 if we set $k(t) = \sum_{k=1}^{\infty} \frac{1}{(j+kM)!} t^{j+kM}$ and have M age functions $f^{(j)}(x) = \left.\frac{d^j g}{dy^j}\right|_{y=-x}$ present in the model. It is interesting to note, therefore, that transformations of this form do not involve the static age function, as there is no term in the Taylor expansion of $g(t-x)$ corresponding to $a(x)$.[19]

Equation 4.16 has solutions of the form

$$g(y) = \sum_{i=1}^{M} \Re[a_i \ \exp(k_i y)]$$

where $\Re[z]$ is the real part of the expression z, and the k_i are the M roots of the equation $k_i^M = K$. In general, these roots will be complex, and, therefore, $g(y)$ will be exponential, trigonometric or a combination of the two. In addition

$$\begin{aligned} f^{(j)}(x) &= \left.\frac{d^j g}{dy^j}\right|_{y=-x} \\ &= \sum_{i=1}^{M} \Re[a_i k_i^j \ \exp(-k_i x)] \end{aligned}$$

and so the age functions present in the model will also be exponential or trigonometric. □

Exponential age/period terms can be included in models constructed using the "general procedure" of Chapter 5, where they are typically used to explain infant mortality. As

[19] This means that they are also present in models without a static age function, as discussed in Appendix 4.B.

an example, consider a model of the form

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + e^{-\lambda x}\kappa_t^{(2)} + \gamma_{t-x} \tag{4.17}$$

This is an extension of the "exponential" model of Chapter 3, with an additional cohort term. We typically require $\lambda > 0$ to give the age function the demographic significance of governing rates of mortality at low ages. This model will allow the parameters to be transformed using

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\alpha_x, \kappa_t^{(1)}, \kappa_t^{(2)} - a\ e^{\lambda t}, \gamma_y + a\ e^{\lambda y}\} \tag{4.18}$$

This means that exponential trends in time within the (transformed) data are not uniquely identifiable as either age/period or cohort effects.[20] This transformation gives us an extra degree of freedom in the model which could be used to impose an additional identifiability constraint.

In this case, however, the imposition of an identifiability constraint will be of little practical importance. In Section 4.3, we said that in order to be practically important, the unidentifiable deterministic trends must be present in both the age and period dimensions of the transformed data. Whilst exponentially increasing trends in the age function are frequently observed in the data (due to low age mortality effects), exponential trends in the period functions are not.[21] We therefore do not experience problems when fitting the model to data as a result of any failure to be able to assign uniquely such a trend to the either age/period or the cohort terms.

As another example, consider a model with trigonometric age functions of the form

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \cos(\theta x)\kappa_t^{(2)} + \sin(\theta x)\kappa_t^{(3)} + \gamma_{t-x} \tag{4.19}$$

For this model, we can transform the parameters using

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\kappa}_t^{(3)}, \hat{\gamma}_y\} = \{&\alpha_x, \kappa_t^{(1)},\\ &\kappa_t^{(2)} - a\ \cos(\theta t) - b\ \sin(\theta t),\\ &\kappa_t^{(3)} + a\ \sin(\theta t) + b\ \cos(\theta t),\\ &\gamma_y + a\ \cos(\theta y) + b\ \sin(\theta y)\}\end{aligned} \tag{4.20}$$

[20] Note that this transformation has $g(y) = a\exp(\lambda y)$ and therefore $\frac{dg}{dy} = \lambda g$ as per Equation 4.16.

[21] An exponential increase or decrease in the period function will typically correspond to super-exponential growth or decline in the observed mortality rates if either $\eta_{x,t} = \ln(\mu_{x,t})$ or $\eta_{x,t} = \text{logit}(q_{x,t})$. Super-exponential growth in mortality rates are not typically observed.

This means that periodic patterns are not uniquely identifiable as either age/period or cohort effects.[22]

As with the exponential functions, the presence of unidentifiable trigonometric trends in the model will be of little practical importance. Whilst the (transformed) data often exhibits periodic behaviour in the cohort and period effects, it is rare to see periodic behaviour across ages.[23] Again, we do not have the unidentifiable deterministic trends for the model in both the age and period dimensions and consequently do not experience practical difficulties when fitting the model to data as a result of any failure to be able to assign uniquely such trends to the either age/period or the cohort terms.

4.4.3 Other age functions

Other parametric age functions do not admit any additional invariant transformations involving the cohort parameters, except in the case where they are actually redefined polynomials, exponentials or trigonometric functions. For instance, the third age/period term in the Plat model did not generate any extra interactions with the cohort parameters, beyond those of the reduced Plat model. This simplifies the identifiability issues of more complex mortality models with different types of age functions, such as those produced by the "general procedure" of Chapter 5, compared with what would otherwise be necessary, were, for instance, only polynomial age functions to be used.

4.4.4 Summary

In summary, issues with the identifiability of APC models relate to functions of year of birth which can be decomposed into purely age/period terms. However, this is only true in models where the age functions take specific parametric forms - namely polynomial, exponential and trigonometric functions. In such models, certain deterministic trends cannot be uniquely allocated between the age/period and cohort terms in the model and so require the imposition of arbitrary identifiability constraints in order to uniquely specify the model.[24] This is summarised in the flow chart in Figure 4.1.

[22] Note that this transformation has $g(y) = a\ \cos(\theta y) + b\ \sin(\theta y)$ and therefore $\frac{d^2 g}{dy^2} = -\theta^2 g$ as per Equation 4.16.

[23] The lack of periodic structure across ages also explains why trigonometric age functions are not widely used in practice.

[24] As discussed in Appendix 4.B, APC mortality models with non-parametric age functions will not have any additional transformations that leave the fitted mortality rates exactly unchanged. However,

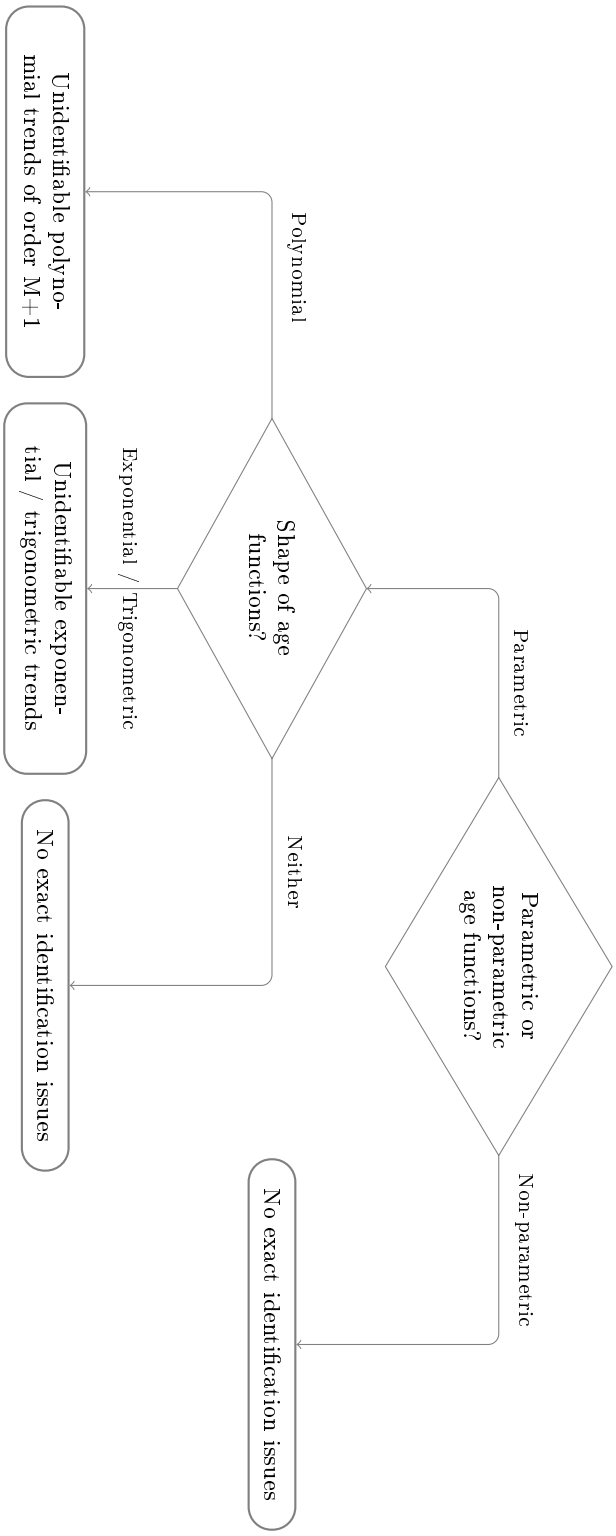


FIGURE 4.1: Flow chart of identifiability issues in APC models

4.5 Projection

In the preceding sections, we have seen that APC mortality models are not fully identified and that we can impose arbitrary identifiability constraints on the parameters in order to fit them to the historical data. Two different modellers using the same data and the same model but different arbitrary identification constraints will obtain different sets of parameters, but these will give identical fitted mortality surfaces and, therefore, fits to the data.

For the majority of practical purposes, we not only need to fit a mortality model to historical data but also to use it to project mortality rates into the future. In Chapter 3, we found that we needed to be careful when doing so in AP mortality models in order to ensure that the projected mortality rates will not depend on the arbitrary identifiability constraints imposed when fitting the models to data. The same is true in APC mortality models. However, the addition of a set of cohort parameters and the presence of unidentifiable deterministic trends complicate this analysis significantly.

The most obvious change when moving from an AP to an otherwise similar APC mortality model is the presence of a set of cohort parameters which will also need to be projected into the future. The period and cohort parameters in the APC model are conceptually different and need to be treated separately when making projections. This is because cohort effects have very different demographic significance from the period effects and are treated separately when fitting the model. It is therefore common practice to project the period and cohort parameters independently.

Some authors (e.g., Haberman and Renshaw (2011)) disagree with this approach, arguing that it may only be appropriate to do this when the cohort parameters are estimated using the residuals from the fitted primary age/period structure. This means that the cohort structure fitted by the model is independent of the age/period structure by construction. However, such fitting techniques will not give parameter estimates which maximise the fit to data and can lead to hierarchical issues (because the cohort parameters are only estimated conditional on the previously fitted estimates of the age/period structure). We, therefore, have a clear preference for model fitting techniques where all

such models may have transformations that leave the fitted mortality rates approximately unchanged, as discussed in Hunt and Villegas (2015).

parameters are estimated together in order to generate the best fit to the historical data.[25]

More generally, it is conceivable that events such as influenza pandemics will cause both an immediate rise in mortality and also lifelong health effects in infants born during the pandemic due to selection effects, leading to correlations between extreme period and cohort effects. However, it is difficult to analyse any dependence structure between the cohort and period parameters as the cohort parameters will be observed over a longer time period, but potentially at a lag of some decades. While it is possible that some extreme mortality events may generate distinctive effects in both the period and cohort parameters, the evidence supporting this conjecture is currently ambiguous (for instance, see Murphy (2009)) and will not generally be relevant for more typical period and cohort effects. An assumption of independence is, therefore, both practical and parsimonious.

In order to make projections of future mortality rates, we typically model the period and cohort parameters as being generated by independent time series processes and use these to project the parameters stochastically into the future. However, the precise form of the time series processes generating the parameters is unknown. Therefore, we analyse the fitted parameters by statistical methods, such as the Box-Jenkins procedure, to determine which processes from the ARIMA family provide the best fit.

Nevertheless, when it comes to projecting mortality rates, we need to recognise that there is a fundamental symmetry between the processes of estimating a model and projecting it: the former takes observations to calibrate the model, whilst the latter uses this calibration to produce projected observations of the future. Due to this symmetry, identification issues which exist when fitting the model may also yield problems when projecting it. When estimating the model, these identifiability issues were solved by imposing arbitrary identifiability constraints on the parameters. However, any time series structure that we find in the parameters needs to be independent of the arbitrary identification scheme used when fitting the model to historical data.

We formalise this by saying that:

> Two sets of model parameters, which give identical fitted mortality rates for the past, should give identical projected mortality rates when projected into the future.

[25] For example, in the general procedure of Chapter 5, all parameters are re-estimated every time the structure of the model is changed, in order to ensure a close fit to the data.

We say that time series processes which satisfy this property are "well-identified".

In particular, the invariant transformations of the parameters of the model which leave the fitted mortality rates unchanged should also leave the projected mortality rates unchanged and, hence, the time series processes used to generate the projected mortality rates unchanged. Consequently, we should use the same time series processes for all sets of parameters from a model which give the same fitted mortality rates. If this is not the case, different processes will be used for different arbitrary identifiability constraints, giving different projected mortality rates. A well-identified time series process should be equally appropriate for all equivalent sets of parameters. To confirm this, we need to check that applying the invariant transformations to the parameters, which leave the fitted mortality rates unchanged, do not also affect the time series processes used to project the parameters.

Chapter 3 discussed how the identification issues in the class of AP models meant that methods for projecting the period parameters from these models into the future needed to be chosen with care in order to ensure they are well-identified. In general, we argued that we should choose to project the model using multivariate methods which are as unstructured as possible, i.e., we should not impose features such as independence, levels of mean reversion or different orders of integration on the time series a priori, but allow these to emerge during the fitting process. However, we also saw that, in models with parametric age functions, the age/period terms were no longer interchangeable once we defined their forms in the model. This allowed us to prioritise biological reasonableness[26] over using the same processes for equivalent models, i.e., models giving the same fitted mortality rates with different definitions of the age functions.

Current practice is to:

1. fit the chosen model to data, imposing any arbitrary identifiability constraints needed in order to specify the parameters uniquely;

2. select time series processes for projecting the parameters based on either using a statistical method (such as the Box-Jenkins procedure to select the preferred processes from the ARIMA class of models) or by directly choosing the time series processes to ensure biologically reasonable projections by making an appeal to the demographic significance of the parameters.

[26] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

However, such an approach often leads to projections of mortality rates which are not well-identified. This is because the second step assumes that the parameters found at the first step are known, rather than merely estimated up to an arbitrary identifiability constraint. This means that current practice builds the arbitrary identifiability constraint into the projection process, ensuring that the projected mortality rates are also arbitrary.

To avoid this, we propose to work backwards from our desire for projections which are biologically reasonable and well-identified to determine the time series processes we need to use to achieve these aims. Before fitting the model, we need to conduct a thorough analysis of the identifiability issues in the chosen model, using the principles established in Section 4.4, to determine which features of the parameters are set by the data and which are set by the arbitrary identifiability constraints. Then, suitable time series processes should be selected to model only the former, identifiable features of the parameters, while still allowing for the unidentifiable trends in a way that guarantees that they do not affect the projection of future mortality rates. By following this procedure, we can ensure that the time series processes are well-identified and that the projected mortality rates do not depend on the arbitrary choices we make when fitting the model.

In this section, we will first look at the broad set of criteria needed for well-identified projection methods in general APC mortality models in Section 4.5.1. Section 4.5.2 looks in more detail at why current practice can lead to projections which are not well-identified and depend on the arbitrary identifiability constraints chosen in the context of the classic APC model from Section 4.3. We then revisit the general case of an APC mortality model in Section 4.5.3, in order to determine general rules for choosing time series processes which are well-identified. These are then applied in the context of the classic APC model again in Section 4.5.4 and it is demonstrated that projected mortality rates are genuinely independent of the choice of arbitrary identifiability constraint. Section 4.5.5 then applies the general rules in the context of the Plat model from Plat (2009a) and Section 4.4.1.1 to see how they work in the context of more sophisticated mortality models with more complex identifiability issues.

4.5.1 Projecting general APC models

Consider the case of projecting an APC mortality model, which has been fitted using data over the period $[1, T]$ to give mortality rates at time $\tau > T$. From Equation 4.2, we

could write this as

$$\eta_{x,\tau} = \alpha_x + \boldsymbol{\beta}_x^\top \boldsymbol{\kappa}_\tau + \gamma_{\tau-x}$$

If the model has identifiability issues, then the projected mortality rates should be unchanged under exactly the same invariant transformations as the fitted mortality rates were, i.e., if we have an invariant transformation of the form of Equation 4.8, namely

$$\begin{aligned}
\hat{\alpha}_x &= \alpha_x - a(x) \\
\hat{\boldsymbol{\beta}}_x &= \boldsymbol{\beta}_x \\
\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\kappa}_t - \boldsymbol{k}(t) \\
\hat{\gamma}_y &= \gamma_y + g(y)
\end{aligned}$$

where $a(x)$, $k^{(i)}(t)$ and $g(y)$ satisfy Equation 4.9, in which case

$$\eta_{x,\tau} = \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top \hat{\boldsymbol{\kappa}}_\tau + \hat{\gamma}_{\tau-x}$$

The projected $\boldsymbol{\kappa}_\tau$ (and potentially the $\gamma_{\tau-x}$) will be random variables, whose distribution is a function of the historical, fitted values, i.e., $\boldsymbol{\kappa}_\tau = P_\kappa(\tau; \{\boldsymbol{\kappa}\})$ and $\gamma_y = P_\gamma(y; \{\gamma\})$. We said previously that we should use the same method of projection for all sets of parameters as a first step to ensure that the projected mortality rates do not depend upon the identifiability constraints. However, for different identifiability constraints, these processes will be estimated from different sets of fitted parameters, e.g., if we use $P_\kappa(\tau; \{\boldsymbol{\kappa}\})$ to project the untransformed period parameters, we must use $P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\})$ to project the transformed period parameters. If we combine this with the invariance of the projected mortality rates, we have

$$\begin{aligned}
\alpha_x + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa}\}) + P_\gamma(\tau - x; \{\gamma\}) &= \hat{\alpha}_x + \hat{\boldsymbol{\beta}}_x^\top P_\kappa(\tau; \{\hat{\boldsymbol{\kappa}}\}) + P_\gamma(\tau - x; \{\hat{\gamma}\}) \\
&= \alpha_x - a(x) + \boldsymbol{\beta}_x^\top P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) + P_\gamma(\tau - x; \{\gamma + g\}) \\
P_\gamma(\tau - x; \{\gamma + g\}) - P_\gamma(\tau - x; \{\gamma\}) &= a(x) + \boldsymbol{\beta}_x^\top \left(P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\})\right)
\end{aligned}$$

Using Equation 4.9, we can eliminate $a(x)$

$$P_\gamma(\tau - x; \{\gamma + g\}) - P_\gamma(\tau - x; \{\gamma\}) = g(\tau - x) + \boldsymbol{\beta}_x^\top \left(P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) - \boldsymbol{k}(\tau)\right)$$

In order for this to hold for all τ and x requires

$$P_\kappa(\tau; \{\boldsymbol{\kappa} - \boldsymbol{k}\}) = P_\kappa(\tau; \{\boldsymbol{\kappa}\}) - \boldsymbol{k}(\tau) \tag{4.21}$$

$$P_\gamma(y; \{\gamma + g\}) = P_\gamma(y; \{\gamma\}) + g(y) \tag{4.22}$$

This means that we should obtain the same results if we project the transformed parameters as if we transform the projected parameters, i.e., the processes of projection and transformation are commutative. Consequently, we see that, in order for a projection method to be well-identified under the invariant transformation, it needs to preserve the unidentifiable trends in the model, i.e., P_κ must preserve the trends $\boldsymbol{k}(t)$, and P_γ must preserve the trend $g(y)$. This also means that it does not matter in which order we perform the processes of projection and transformation, the distribution of the transformed parameters projected into the future will be identical to the distribution of the projected parameters which are then transformed.

In addition, since

$$\mathbb{V}ar(\boldsymbol{\kappa}_\tau) = \mathbb{V}ar(\boldsymbol{\kappa}_\tau - \boldsymbol{k}(\tau)) = \mathbb{V}ar(\hat{\boldsymbol{\kappa}}_t)$$
$$\mathbb{V}ar(\gamma_y) = \mathbb{V}ar(\gamma_y + g(y)) = \mathbb{V}ar(\hat{\gamma}_y)$$

we note that the variability of the parameters around the trend is identifiable and so does have a meaning independent of the identifiability constraints imposed. Therefore, we conclude that, while the deterministic trends may be unidentifiable and not meaningful, the variation around the trend is of genuine significance, since it is independent of the identifiability constraints. Therefore, this variation needs to be projected consistent with our demographic significance for the parameters and what has been observed in the historical data.

However, the time series processes selected via current practice often do not preserve the unidentifiable trends in the period and cohort parameters, as we shall now see using the classic APC model.

4.5.2 Projecting the classic APC model

It has long been known, at least since Osmond (1985), that the lack of identifiability in the classic APC model has important consequences when making projections from the model. Different sets of arbitrary identifiability constraints are based on different allocations of the linear trends in the data between the age, period and cohort parameters. The outcome of current practice can therefore be influenced by the presence or absence of a linear trend in the fitted parameters, despite this being purely dependent upon the identifiability constraints chosen.

To illustrate this, we consider projecting the classic APC model fitted using four different sets of identifiability constraints. The fitted mortality rates given using these four sets of constraints are identical; however, the time series processes found by current practice differ which means that current practice would give different projected mortality rates in the four different cases. Consequently, these time series processes are not well-identified.

We start by fitting the classic APC model to mortality data for the USA from Human Mortality Database (2014) for ages 50 to 100 and year 1950 to 2010. As discussed in Section 4.3, a number of equally valid identifiability constraints can be imposed on this model, which give identical fitted mortality rates. We consider the following four sets of identifiability constraints:

Case 1: $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x} = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = \sum_{x,t} \gamma_{t-x}((t - \bar{t}) - (x - \bar{x})) = 0$. This was discussed in Section 4.3 and restricts the cohort parameters to be zero on average and without any linear trends, consistent with our desired demographic significance for the cohort parameters.

Case 2: $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y (y - \bar{y}) = 0$. These constraints impose the same demographic interpretation on the parameters, except that the averages are not weighted by the number of observations of each cohort.

Case 3: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(x - \bar{x}) = 0$. This set of constraints is the same as imposed on the classic APC model in Cairns et al. (2009), where it was written as imposing $\sum_x (\alpha_x - \frac{1}{T}\sum_t \eta_{x,t})(x - \bar{x}) = 0$, i.e., that the static age function, α_x, explains all the linearity across ages in the data.

Case 4: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(t - \bar{t}) = 0$. Similar to Case 3, this set of constraints imposes that the period function, κ_t, accounts for all of the linearity across years in the data.

The first thing to note is that all of these constraints were developed to give the cohort parameters the same demographic significance, i.e., that they should be centred on zero and the other functions in the model should capture any linear trends. Because of this, the fitted parameters in each case are very similar. However, they are not identical, unlike the fitted mortality rates. We therefore see that demographic significance, whilst helpful in selecting an appropriate set of identifiability constraints, does not specify a unique set of constraints to use. Model users with the same interpretation of the parameters can reasonably choose to impose different constraints and obtain different fitted parameters when using the same model with the same data. The fact that demographic significance is subjective and, in practice, different model users adopt a range of interpretations for

the different parameters highlights the fact that we must take care to ensure that any conclusions regarding projected mortality rates are independent of the arbitrary choice of constraints made when fitting the model, and underscores the extent to which the identifiability constraints we choose is arbitrary.

Current practice is to take the fitted parameters and then determine which time series processes to use to project them. This may involve performing a Box-Jenkins analysis on the fitted parameters, as was done in Lee and Carter (1992) and Cairns et al. (2011a). Alternatively, current practice may appeal to the demographic significance assigned to the parameters, as in Plat (2009a). Such an appeal might determine that the period function is non-stationary (as it is primarily responsible for the evolution of mortality) and, based on the discussion in Chapter 2, that the cohort parameters are stationary around zero. It might therefore appear reasonable to choose[27] to use a random walk with drift process for κ_t and an AR(1) process for γ_y

$$\kappa_t = \kappa_{t-1} + \mu + \epsilon_t \tag{4.23}$$

$$\gamma_y = \rho\gamma_{y-1} + \varepsilon_y \tag{4.24}$$

Table 4.1 shows the fitted parameters for the four cases above using these time series processes.

	Case 1	Case 2	Case 3	Case 4
κ_{2010}	-0.3526	-0.3439	-0.3550	-0.3478
μ	-0.0110	-0.0107	-0.0111	-0.0109
$\sigma_\kappa = \mathbb{S}t\mathbb{D}ev(\epsilon_t)$	0.0161	0.0161	0.0161	0.0161
γ_{1950}	-0.1459	-0.1125	-0.1422	-0.1530
ρ	0.9513	0.9577	0.9499	0.9542
$\sigma_\gamma = \mathbb{S}t\mathbb{D}ev(\varepsilon_y)$	0.0193	0.0184	0.0193	0.0194

TABLE 4.1: Time series parameters for the period and cohort functions in the classic APC model fitted using different identifiability constraints

For $\tau - x > 1950$,[28] we find

$$\mathbb{E}\eta_{x,\tau} = \alpha_x + \kappa_{2010} + (\tau - 2010)\mu + \rho^{\tau-x-1950}\gamma_{1950} \tag{4.25}$$

[27] Note that we are not saying that these are the most appropriate time series processes to use for this set of parameters. We use them for illustrative purposes as they are relatively simple and not atypical of the processes used in practice. However, it is important to observe that selecting alternative time series processes on a purely statistical basis from the fitted parameters would not solve the issues we have identified.

[28] That is, for cohort parameters that are projected rather than fitted from historical data, taking into consideration that cohort parameters for the ten most recent years of birth are not fitted from the data due to insufficient observations.

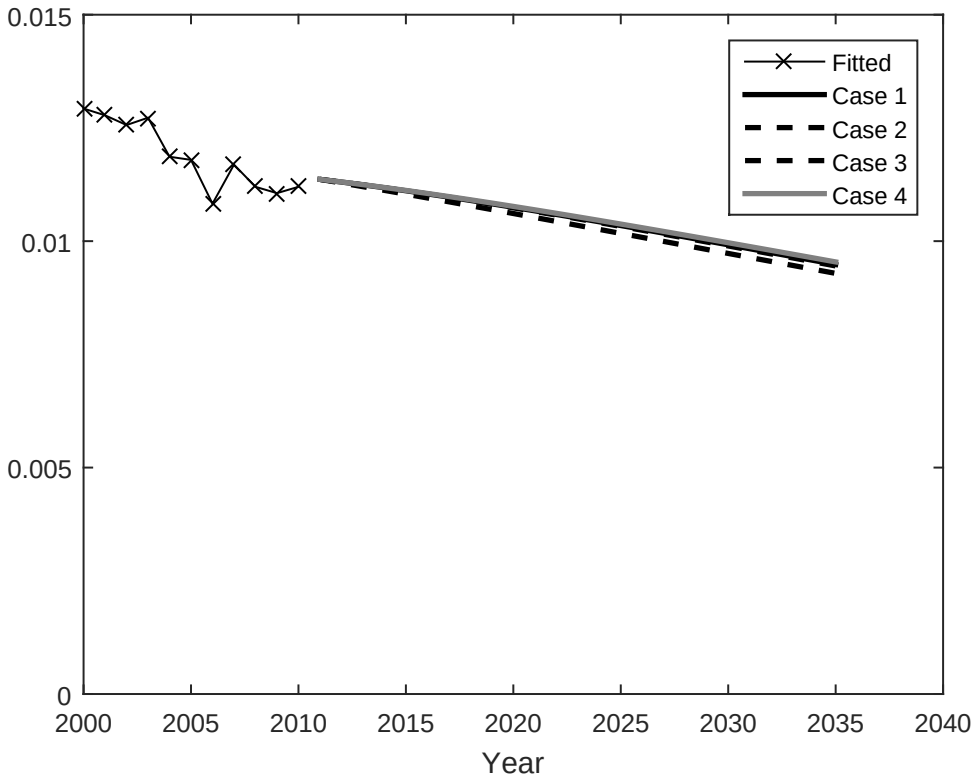


FIGURE 4.2: Projected $\mu_{60,t}$ using different sets of identifiability constraints

We can therefore see that, inserting the fitted time series parameters from Table 4.1 for the four different cases, we do not find the same expected values for the future mortality rates.[29] This is shown in Figure 4.2. In addition, the variability of the projected parameters depends on σ_κ, ρ and σ_γ. However, ρ and σ_γ differ between cases, meaning that the variability of projected mortality rates will also be different for the different cases. These differences in the distribution of projected mortality rates might be felt to be relatively small, although they will grow with projection time. However, the most important point is that the differences should not exist at all - the fitted mortality rates for the different cases were *identical* and so should be the distribution of the projected mortality rates. We therefore see that the time series processes used above to project the classic APC model are not well-identified.

4.5.3 Projecting general APC mortality models: Revisited

From Section 4.5.1 above, we note that we must use the same time series processes to project sets of parameters which give identical fitted mortality rates, i.e., if $P_\gamma(y;\{\gamma\})$

[29] For example, $\mathbb{E}\eta_{60,2020} = -4.5449$ for the Case 1 parameters, -4.5598 for the Case 2 parameters, -4.5459 for the Case 3 parameters and -4.5433 for the Case 4 parameters.

is a suitable process (with time series parameters estimated from the fitted cohort parameters, $\{\gamma_y\}$), then $P_\gamma(y;\{\hat{\gamma}\})$ is a suitable process, albeit with time series parameters estimated from the transformed cohort parameters, $\{\hat{\gamma}_y = \gamma_y + g(y)\}$.

In practice, we usually describe our projection methods in terms of time series processes rather than projection functions. However, the two are equivalent, since the projection function is found by "solving" the difference equation form of the time series. For instance, the AR(1) process has the difference equation form in Equation 4.24, but has solution

$$P_\gamma(y;\{\gamma\}) = \rho^{y-Y}\gamma_Y + \sum_{s=Y+1}^{y} \rho^{y-s}\varepsilon_s$$

where Y is the last year of birth for which we fitted the cohort parameters.

The general form of ARIMA difference equations for γ_y can be written as[30]

$$(1-L)^d\Phi(L)(\gamma_y - \Gamma(y)) = \Psi(L)\varepsilon_y \tag{4.26}$$

where L is the lag operator, d is the order of integration of the process, Φ and Ψ are polynomials of order p and q governing the autoregressive and moving average parts of the process, respectively,[31] ε_y are the innovations and $\Gamma(y)$ is a deterministic function of year of birth. Taking unconditional expectations (i.e., with no conditioning on previous lags of the process), we see that

$$\mathbb{E}\left[\gamma_y - \Gamma(y)\right] = 0 \quad \forall y$$

and that the function $\Gamma(y)$ represents the trend around which the cohort parameters vary.

The invariant transformation of the model in Equation 4.9 adds a deterministic function - the unidentifiable trend $g(y)$ - to the cohort parameters. However, this deterministic function must not change the error term, ε_y, of a well-identified process and so

$$\begin{aligned}
\varepsilon_y &= (1-L)^d\Psi^{-1}(L)\Phi(L)(\gamma_y - \Gamma(y)) \\
&= (1-L)^d\Psi^{-1}(L)\Phi(L)(\hat{\gamma}_y - \hat{\Gamma}(y)) \\
&= (1-L)^d\Psi^{-1}(L)\Phi(L)(\gamma_y + g(y) - \hat{\Gamma}(y))
\end{aligned}$$

[30] For simplicity, we use the cohort function as an illustrative case. The analysis is identical for $\boldsymbol{\kappa}_t$, however.

[31] In order to be stationary, these polynomials have roots with modulus less than unity.

In order to ensure that the variation around the trend, given by the error term, remains unchanged by the invariant transformation, we require

$$\hat{\Gamma}(y) = \Gamma(y) + g(y)$$

In this case, the deterministic trend, $\Gamma(y)$, has changed under the invariant transformation but not the variation around the trend.

We stated above that the time series processes being used for the parameters should be equally applicable for all sets of parameters which give the same fitted mortality rates. This implies that the form of the deterministic trends should be the same, and, therefore, that $\hat{\Gamma}(y)$ is of the same form as $\Gamma(y)$. This can only be true if $\hat{\Gamma}(y)$, $\Gamma(y)$ and $g(y)$ are all of the same form. For instance, if $g(y)$ is a linear function of year of birth (as in the case of the classic APC model), then $\Gamma(y)$ and $\hat{\Gamma}(y)$ must also be linear functions of year of birth and so will not change form under the invariant transformations of the model.

If we solve Equation 4.26, we see that

$$\gamma_y = P_\gamma(y; \{\gamma\}) = \frac{\Psi(L)}{(1-L)^d \Phi(L)} \varepsilon_y + \Gamma(y) \tag{4.27}$$

In this form, it can also be seen that such time series processes preserve unidentified trends in the manner discussed in Section 4.5.1

$$\begin{aligned} \hat{\gamma}_y &= \gamma_y + g(y) \\ &= \frac{\Psi(L)}{(1-L)^d \Phi(L)} \varepsilon_y + \Gamma(y) + g(y) \\ &= \frac{\Psi(L)}{(1-L)^d \Phi(L)} \varepsilon_y + \hat{\Gamma}(y) \end{aligned}$$

i.e., the projected parameters after applying the invariant transformation will have the same variation, $\frac{\Psi(L)}{(1-L)^d \Phi(L)} \varepsilon_y$, but around a different deterministic trend, $\hat{\Gamma}(y)$, compared with the original parameters projected using the same method. The use of the invariant transformations will not affect our measurement of any coefficients in $\Psi(L)$ or $\Psi(L)$ at the fitting stage. Thus, we also see that the two ways of looking at the projected parameters, namely as time series processes and via projection functions, are equivalent.

As an example, consider the cohort parameters in the classic APC model. From Section 4.3, we see that, in this model, the cohort parameters have an unidentified constant and linear trend, i.e., $g(y) = b + c(y - \bar{y})$ from Equations 4.5 and 4.6. In Section 4.5.2, we

said that current practice might use an AR(1) process for the cohort parameters, which has ARIMA form

$$(1-\rho L)\gamma_y = \varepsilon_y$$

Comparing this with Equation 4.26, we see that current practice assumes that $\Gamma(y) = 0$, which is not of the same form as $g(y)$ above. Therefore, the time series process changes form when using an alternative set of parameters $\hat{\gamma}_y = \gamma_y + g(y)$ in place of γ_y,

$$\begin{aligned}(1-\rho L)\hat{\gamma}_y &= (1-\rho L)(\gamma_y + b + c(y-\bar{y})) \\ &= (1-\rho L)\gamma_y + (1-\rho)(b + c(y-\bar{y})) + \rho c \\ &= \varepsilon_y + (1-\rho)(b + c(y-\bar{y})) + \rho c \\ &\neq \varepsilon_y\end{aligned}$$

and therefore the process is not well-identified.

When analysed in this form, however, a solution becomes immediately apparent: we need to introduce a linear function, $\Gamma(y) = \beta_0 + \beta_1 y$, into the AR(1) process to ensure that the process is well-identified, i.e.,

$$(1-\rho L)(\gamma_y - \beta_0 - \beta_1 y) = \varepsilon_y \tag{4.28}$$

Using the alternative parameters $\hat{\gamma}_y$ would produce

$$\begin{aligned}(1-\rho L)(\hat{\gamma}_y - \hat{\beta}_0 - \hat{\beta}_1 y) &= (1-\rho L)(\gamma_y + b + c(y-\bar{y}) - \hat{\beta}_0 - \hat{\beta}_1 y) \\ &= (1-\rho L)(\gamma_y - \beta_0 - \beta_1 y) \\ &= \varepsilon_y\end{aligned}$$

if $\hat{\beta}_0 = \beta_0 - b - c\bar{y}$ and $\hat{\beta}_1 = \beta_1 - c$. Therefore, the form of Equation 4.28 does not change under the invariant transformations of the classic APC model, and we conclude that this time series process is well-identified. Again, we also see that the variation around the linear trend, given by ε_y, is unchanged by the invariant transformation, whilst the unidentifiable trend is affected by the invariant transformation.

The time series process in Equation 4.28 has been suggested previously for the cohort parameters in Cairns et al. (2009) where it was referred to as the "*AR(1) process around a linear drift*". However, in Cairns et al. (2009), it was not used for the classic APC model, nor was it selected for being well-identified, but rather on the grounds of fitting

the observed cohort parameters well.

The AR(1) around linear drift process is solved to give

$$P_\gamma(y;\{\gamma\}) = \rho^{y-Y}(\gamma_Y - \beta_0 - \beta_1 Y) + \beta_0 + \beta_1 y + \sum_{s=Y+1}^{y} \rho^{y-s}\varepsilon_s$$

We can also verify, by substituting the forms for $\hat{\gamma}_y$, $\hat{\beta}_0$ and $\hat{\beta}_1$ found above, that this process also satisfies the requirement of Equation 4.22 in Section 4.5.1, namely

$$P_\gamma(y;\{\hat{\gamma}\}) = P_\gamma(y;\{\gamma\}) + a + b(y - \bar{y})$$

Hence, projecting the transformed cohort parameters gives us the same results as transforming the projected cohort parameters.

Returning to the form of the time series process in Equation 4.26, it is common to write this in an alternative, but equivalent form

$$\begin{aligned}(1-L)^d\Phi(L)\gamma_y - (1-L)^d\Phi(L)\Gamma(y) &= \Psi(L)\varepsilon_y \\ (1-L)^d\Phi(L)\gamma_y &= \xi(y) + \Psi(L)\varepsilon_y \end{aligned} \tag{4.29}$$

where $\xi(y)$ is a deterministic function of y and $\Gamma(y)$ solves the difference equation

$$(1-L)^d\Phi(L)\Gamma(y) = \xi(y) \tag{4.30}$$

In this form, $\xi(y)$ is often referred to as the "drift". Knowing the form that $\Gamma(y)$ must take (i.e., the same form as $g(y)$ from the unidentifiable trends in the model in Equation 4.8), we can therefore specify the correct form of $\xi(y)$.

As an example of this, consider the classic APC model again, but, this time, consider the period parameters. We know from Section 4.3 that the period parameters have an unidentified linear trend in much the same way as the cohort parameters, i.e., $k(t) = a - c(t - \bar{t})$ if we re-write Equations 4.4 and 4.6 using the notation of Equation 4.9. Random walk processes are often used for the period parameters, i.e., we assume $d = 1$ and $\Phi(L) = \Psi(L) = 1$. It is then important to specify the correct form for the drift $\xi(t)$. Based on similar arguments to the ones used above for the cohort parameters, we should look for time series processes of the form

$$(1-L)(\kappa_t - \nu_0 - \nu_1 t) = \epsilon_t$$

which has a linear trend $K(t) = \nu_0 + \nu_1 t$. To obtain a well-identified time series of the form of Equation 4.29, we need the drift, $\xi(t)$, of the random walk to satisfy

$$\begin{aligned}\xi(t) &= (1-L)(\nu_0 + \nu_1 t) \\ &= \nu_0 + \nu_1 t - \nu_0 - \nu_1(t-1) \\ &= \nu_1\end{aligned}$$

i.e., the drift is constant. This shows that the random walk with drift is well-identified for the period parameters in the classic APC model.

We can also verify this directly, since

$$\begin{aligned}\epsilon_t &= \kappa_t - \kappa_{t-1} - \mu \\ &= \hat{\kappa}_t - a + c(t-\bar{t}) - \hat{\kappa}_{t-1} + a - c(t-1-\bar{t}) - \mu \\ &= \hat{\kappa}_t - \hat{\kappa}_{t-1} - \hat{\mu}\end{aligned}$$

if $\hat{\mu} = \mu - c$. Thus the transformed period parameters, $\hat{\kappa}_t$, follow a random walk with drift if the original period parameters do. However, the value of the drift, which determines the unidentifiable linear trend, will change under the invariant transformation, although the innovations, ϵ_t, which determine the variability around this drift do not.

In summary, we have the following procedure for selecting a well-identified time series process for any specific APC mortality model:

1. Determine the identifiability issues in the specific APC model by finding the unidentifiable deterministic trends for the parameters which cannot be assigned between the different age/period and cohort terms in the specific model. This will need to be done prior to the fitting stage in order to fit the model robustly to data.

2. Specify a time series process for the variation around these trends. This can either be done by analysing this variation using statistical techniques, or by selecting a process which accords with our demographic significance for the parameters. Doing so will set the form of $\Phi(L)$ and $\Psi(L)$, which determine the stochastic structure of the ARIMA process.

3. Specify the deterministic trends, $\Gamma(y)$, in the time series process in Equation 4.26, which will need to be of the same form as $g(y)$. Equivalently, this can be achieved by finding a drift function, $\xi(y)$, in the alternative form of the time series process in Equation 4.29, with the requirement that $(1-L)^d \Phi(L)\Gamma(y) = \xi(y)$.

It is important to recognise that this procedure works backwards from the variation around the trends in the parameters, which is independent of the identifiability constraints and then adds back in the unidentifiable trends which will depend upon the specific set of identifiability constraints we use when fitting the model. In this fashion, we can ensure that the projected parameters are both well-identified and possess our desired demographic significance when specifying a suitable form for the time series process.

4.5.4 Projecting the classic APC model: Revisited

In Section 4.5.2, it was demonstrated that the current practice approach to selecting time series processes for the period and cohort parameters in the classic APC model yielded projections of mortality rates which depended upon arbitrary choices made when fitting the model. In Section 4.5.3, we then showed that the issue in this case was not the use of the random walk with drift for the period parameters, but the selection of an AR(1) process, rather than an AR(1) process around a linear drift for the cohort parameters.

If we use the AR(1) around linear drift process for the cohort parameters for the four cases discussed in Section 4.5.2, we obtain the time series parameters in Table 4.2.

	Case 1	Case 2	Case 3	Case 4
γ_{1950}	-0.1459	-0.1125	-0.1422	-0.1530
β_0	0.1388	0.1852	0.1388	0.1388
β_1	-0.0053	-0.0056	-0.0052	-0.0055
ρ	0.9636	0.9636	0.9636	0.9636
$\sigma_\gamma = \mathbb{S}t\mathbb{D}ev(\varepsilon_y)$	0.0184	0.0184	0.0184	0.0184

TABLE 4.2: Time series parameters for different identifiability constraints

As previously mentioned in Section 4.5.2, ρ and σ_γ control the variation of projected cohort parameters. It is, consequently, important to see that these parameters do not change in the four different cases using the well-identified time series processes. The variability of projected mortality rates will be identical in each of the four cases. Using the AR(1) around linear drift process, we also find

$$\begin{aligned}\mathbb{E}\eta_{x,\tau} &= \alpha_x + \kappa_{2010} + (\tau - 2010)\mu \\ &\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) + \beta_0 + \beta_1 \times (\tau - x) \end{aligned} \tag{4.31}$$

From the results of Section 4.5.3, we can see that if we transform the parameters of the classic APC model using the transformation in Equations 4.4, 4.5 and 4.6, and then

project them using well-identified time series processes, we obtain

$$\begin{aligned}
\hat{\alpha}_x &= \alpha_x - a - b + c(x - \bar{x}) \\
\mathbb{E}\hat{\kappa}_\tau &= \hat{\kappa}_{2010} + \hat{\mu}(\tau - 2010) \\
&= \kappa_{2010} + a - c(2010 - \bar{t}) + (\mu - c)(\tau - 2010) \\
&= \kappa_{2010} + a - c(\tau - \bar{t}) + \mu(\tau - 2010) \\
\mathbb{E}\hat{\gamma}_{\tau-x} &= \rho^{\tau-x-1950}(\hat{\gamma}_{1950} - \hat{\beta}_0 - \hat{\beta}_1 \times 1950) + \hat{\beta}_0 + \hat{\beta}_1 \times (\tau - x) \\
&= \rho^{\tau-x-1950}(\gamma_{1950} + b + c(1950 - x - \bar{y}) - \beta_0 - b - c\bar{y} - (\beta_1 + c) \times 1950) \\
&\quad + \beta_0 + b + c\bar{y} + (\beta_1 + c) \times (\tau - x) \\
&= \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) \\
&\quad + \beta_0 + \beta_1(\tau - x) + c(\tau - x - \bar{y})
\end{aligned}$$

Hence, the expectation of $\eta_{x,t}$ in Equation 4.31, after applying the invariant transformations, becomes

$$\begin{aligned}
\mathbb{E}\hat{\eta}_{x,\tau} &= \hat{\alpha}_x + \hat{\kappa}_{2010} + (\tau - 2010)\hat{\mu} \\
&\quad + \rho^{\tau-x-1950}(\hat{\gamma}_{1950} - \hat{\beta}_0 - \hat{\beta}_1 \times 1950) + \hat{\beta}_0 + \hat{\beta}_1 \times (\tau - x) \\
&= \alpha_x - a - b + c(x - \bar{x}) + \kappa_{2010} + a - c(\tau - \bar{t}) + \mu(\tau - 2010) \\
&\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) \\
&\quad + \beta_0 + \beta_1(\tau - x) + c(\tau - x - \bar{y}) \\
&= \alpha_x + \kappa_{2010} + (\tau - 2010)\mu \\
&\quad + \rho^{\tau-x-1950}(\gamma_{1950} - \beta_0 - \beta_1 \times 1950) + \beta_0 + \beta_1 \times (\tau - x) \\
&= \mathbb{E}\eta_{x,\tau}
\end{aligned}$$

We can therefore see how changes in the linear drift of the period functions between the different cases cancel with the changes in the linear drift in the cohort functions to give exactly the same expected projected mortality rates in all four cases.[32] We, therefore, see in practice what was derived theoretically in Section 4.5.3, namely that using a random walk with drift process for the period parameters and an AR(1) around linear drift process for the cohort parameters gives well-identified projections for the classic APC model, and so the projected mortality rates which do not depend upon the identifiability constraints imposed.

Projections using an AR(1) process around a linear drift might be felt to conflict with our desired demographic significance for the cohort parameters, i.e., that they should

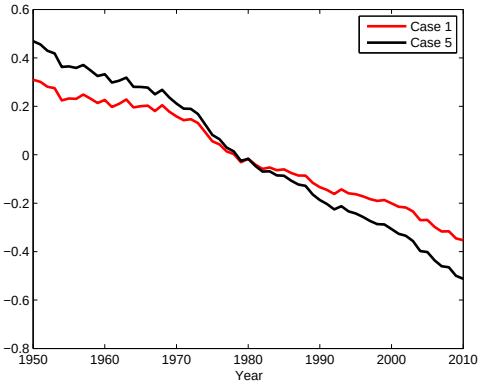
[32] For example, in all four cases $\mathbb{E}\eta_{60,2020} = -4.6413$.

exhibit no long-term trends. However, demographic significance is subjective and so should not be used to override a greater concern that the projected mortality rates do not depend upon the arbitrary identifiability constraints. Fortunately, there are methods for obtaining well-identified projections of the cohort parameters which do conform to our desired demographic significance of trendlessness.

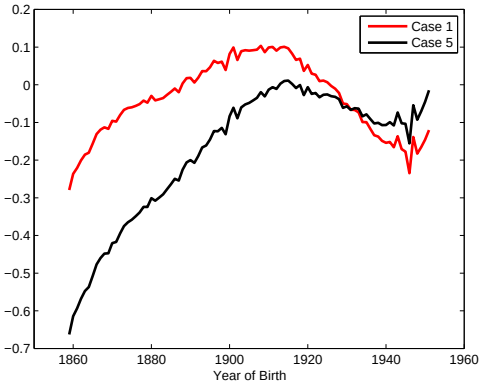
In order to lack trends, the drift coefficients of the process, β_0 and β_1, should be zero. Looking again at Table 4.2, one might think that the values of β_0 and β_1 are quite small, and therefore be tempted to test them statistically with a view to setting them to zero. This, however, would be a mistake. As shown in Section 4.5.3, the values of β_0 and β_1 change under the invariant transformations of the classic APC model and, therefore, will depend upon the identifiability constraints chosen. Consequently, the results of any statistical analysis of their significance will also depend upon the arbitrary identifiability constraints, which is not desirable.

The reason that β_0 and β_1 are "small" is because we have imposed this via the identifiability constraints. All four sets of identifiability constraints were chosen to set the level of the cohort parameters to be around zero and to have no linear trends over the whole range of the data. Therefore, we would expect to find low values of β_0 and β_1, which control the level and drift to which the process mean-reverts. We could have chosen other, equally reasonable constraints based on alternative subjective interpretations of the demographic significance of the period and cohort parameters which would have resulted in far larger values of β_0 and β_1 and given exactly the same fitted and projected mortality rates. We therefore see that whether or not these parameters are "small", and consequently whether or not they pass a statistical test of their significance, is solely dependent upon the arbitrary identifiability constraints we have chosen.

The four cases in Section 4.5.2 were motivated by the same desired demographic significance for the cohort parameters - that they should be centred around zero and not have any linear trends. However, the four different cases used four different interpretations of these subjective requirements, and therefore arrived at four different interpretations of what it means to be centred around zero and trendless. These different interpretations resulted in the four different sets of identifiability constraints. Using an AR(1) around linear drift process to project the cohort functions introduces a fifth interpretation for the meaning of being centred around zero and having no linear drift, in this case, that the time series parameters β_0 and β_1 are equal to zero. Therefore, we could use another set of parameters with the identifiability constraints



(A) Period parameters κ_t



(B) Cohort parameters γ_y

FIGURE 4.3: Projecting the parameters of the classic APC model: Cases 1 and 5

Case 5: $\sum_t \kappa_t = 0$, $\beta_0 = 0$ and $\beta_1 = 0$

This set of constraints gives identical fitted and projected mortality rates to the other cases, but gives projected cohort parameters which mean-revert around zero, which accords better with our demographic significance. However, the restrictions in Case 5 cannot be known at the time of fitting the model to data, since the appropriate time series process that will be used to project the cohort parameters cannot be known at that stage. To use this set of constraints, we need to do the following:

1. fit the model to data, applying some convenient set of identifiability constraints which can be known in advance of analysing the time series structure of the parameters, e.g., those in Case 1;

2. estimate values for β_0 and β_1 for these historical parameters by fitting the AR(1) around a linear drift process in Equation 4.28 to them;

3. use these estimated values for β_0 and β_1 in the transformations in Equations 4.5 and 4.6 to obtain a new set of (equivalent) age, period and cohort parameters.

The period and cohort parameters for Case 5, compared with those for Case 1, are shown in Figure 4.3. Using the Case 5 parameters may appear unnatural as the cohort parameters in this case appear to possess a linear trend. However, when we project using the well-identified AR(1) around linear drift process, we find no linear drift in these parameters, merely mean reversion to a level of zero, which fits well with the demographic significance for the cohort parameters discussed in Chapter 2.

4.5.5 Projecting the Plat model

We will now use this analysis to specify a set of well-identified projection processes for the Plat model discussed in Section 4.4.1.1. As described in that section, the invariant transformations of the model can be written in the form of Equation 4.9 with

$$\begin{aligned}
\hat{\alpha}_x &= \alpha_x - a_1 - a_2 - a_3 - b + c(x-\bar{x}) - d(x-\bar{x})^2 && = \alpha_x - a(x) \\
\hat{\kappa}^{(1)}_t &= \kappa^{(1)}_t + a_1 - c(t-\bar{t}) - d(t-\bar{t})^2 && = \kappa^{(1)}_t - k^{(1)}(t) \\
\hat{\kappa}^{(2)}_t &= \kappa^{(2)}_t + a_2 + 2d(t-\bar{t}) && = \kappa^{(2)}_t - k^{(2)}(t) \\
\hat{\kappa}^{(3)}_t &= \kappa^{(3)}_t + a_3 && = \kappa^{(3)}_t - k^{(3)}(t) \\
\hat{\gamma}_y &= \gamma_y + b + c(y-\bar{y}) + d(y-\bar{y})^2 && = \gamma_y + g(y)
\end{aligned}$$

by composing the transformations in Equations 4.4 (for each period function), 4.5, 4.6 and 4.13.

Starting with the cohort parameters, we may wish to retain the demographic interpretation that they should be stationary and mean reverting and so wish to use an AR(1) structure. However, from the discussion in Section 4.5.3 and the observation that $g(y)$ is quadratic for the Plat model, we therefore require that $\Gamma(y)$ in Equation 4.26 is quadratic. In order to give well-identified projections, we would therefore project the cohort parameters using an AR(1) around quadratic drift process, i.e.,

$$(1-\rho L)(\gamma_y - \beta_0 - \beta_1 y - \beta_2 y^2) = \varepsilon_y \tag{4.32}$$

Simple insertion of $\hat{\gamma}_y = \gamma_y + g(y)$ into this shows that it does not change structure under the invariant transformation and so is well-identified. In principal, we could then decide to switch to an equivalent set of parameters with the constraints $\beta_0 = \beta_1 = \beta_2 = 0$ in the same manner as for the classic APC model. This may be desirable as it gives projected cohort parameters which mean-revert around zero, in line with our demographic significance. In addition, when more complicated methods are used to project the cohort parameters, it might be felt to simplify the process of projection.[33]

For the period parameters, we may wish to use a random walk with drift structure as we did for the classic APC model on the demographic interpretation that the period functions should be non-stationary. This would be written as

$$(1-L)\boldsymbol{\kappa}_t = \boldsymbol{\xi}(t) + \boldsymbol{\epsilon}_t \tag{4.33}$$

[33] For an example where this is the case, see Chapter 6.

where $\boldsymbol{\kappa} = \left(\kappa_t^{(1)}, \quad \kappa_t^{(2)}, \quad \kappa_t^{(3)}\right)^\top$ as discussion in Section 4.2 and similarly for $\boldsymbol{\xi}(t)$ and $\boldsymbol{\epsilon}_t$.

Using this notation, we can group the transformations of the period functions as

$$\begin{aligned}\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\kappa}_t + \begin{pmatrix} a_1 + c\bar{t} - d\bar{t}^2 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} -c + 2d\bar{t} \\ 2d \\ 0 \end{pmatrix} t + \begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix} t^2 \\ &= \boldsymbol{\kappa}_t + \boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2\end{aligned}$$

In Section 4.5.3, we showed that in order to ensure identifiability, we needed

$$\begin{aligned}\boldsymbol{\xi}(t) &= (1-L)(\boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2) \\ &= \boldsymbol{k}_0 + \boldsymbol{k}_1 t + \boldsymbol{k}_2 t^2 - \boldsymbol{k}_0 - \boldsymbol{k}_1(t-1) + \boldsymbol{k}_2(t-1)^2 \\ &= \boldsymbol{k}_1 - \boldsymbol{k}_2 + 2\boldsymbol{k}_2 t \\ &= \begin{pmatrix} -c + 2d\bar{t} + d \\ 2d \\ 0 \end{pmatrix} + 2\begin{pmatrix} -d \\ 0 \\ 0 \end{pmatrix} t\end{aligned}$$

Therefore, we see that, in order for the Plat model to have well-identified projections, we require a constant drift component for $\kappa_t^{(2)}$ (i.e., $\xi^{(2)}(t) = \mu_0^{(2)}$, a constant) and a linear drift component for $\kappa_t^{(1)}$ (i.e., $\xi^{(1)}(t) = \mu_0^{(1)} + \mu_1^{(1)} t$, a linear function of time). This can be written as

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_{t-1} + \mu X_t + \boldsymbol{\epsilon}_t \tag{4.34}$$

where

$$\mu = \begin{pmatrix} \mu_0^{(1)} & \mu_t^{(1)} \\ \mu_0^{(2)} & 0 \\ 0 & 0 \end{pmatrix}$$

and $X_t = \left(1, \quad t\right)^\top$. We can see that this form of the random walk with drift process extends naturally to allow for other unidentifiable trends by choosing the "trend" matrix, X_t, and corresponding "drift" matrix, μ, appropriately. The need to use a random walk with linear drift is often overlooked, for instance in Plat (2009a) and Börger et al. (2013) (who used a model which nests the reduced Plat model) - see also Chapter 6.

We also see that different drifts are required for different period functions in order to give well-identified projections of mortality rates. This runs counter to the desire to

treat all the period functions the same, as discussed in Chapter 3. However, using the same drifts for all the period functions can give projections which are not biologically reasonable. For example, allowing for a quadratic trend in $\kappa_t^{(3)}$ can result in apparent changes in trend which are inconsistent with the historical data. In Chapter 3, we also found that we can treat different period functions differently in models with parametric age functions, because there were no invariant transformations of the model which could be used to interchange the age/period terms. It may, therefore, be preferable to allow for different drifts in different period functions in the Plat (2009a) model to obtain well-identified projected mortality rates which are also biologically reasonable.[34] We should, therefore, be prepared to override the desire to treat the period functions identically if the alternative is to put biological reasonableness at stake. See Chapter 6 for an example of this issue in practice.

4.5.6 Summary

APC mortality models which have unidentifiable trends at the fitting stage require extra care when projected to ensure that the projections do not depend on the identifiability constraints chosen. In general, we find that the projection method used must preserve whatever trends were unidentifiable at the fitting stage. For example, the processes which were well-identified for the classic APC model discussed in Section 4.5.4 preserved linear trends, which were shown to be unidentifiable in Section 4.3.

Such an approach generalises naturally for more complicated mortality models, such as the Plat model discussed in Sections 4.4.1.1 and 4.5.5. However, models with higher order polynomial age functions have higher order unidentifiable trends (as shown in Section 4.4.1), and so require projection processes which allow for these trends. This may cause problems for long term projections.

For example, consider the model

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + ((x - \bar{x})^2 - \sigma_x)\kappa_t^{(3)} + \gamma_{t-x} \tag{4.35}$$

which extends model M7 of Cairns et al. (2009) with a static age function (as was done in Haberman and Renshaw (2011)). We can see that a model of this form possesses age functions which span the polynomials to quadratic order. From Section 4.4.1, we know, without performing any additional analysis, that it has unidentifiable cubic trends

[34] Using different drifts for the different period functions will mean, however, that time series processes will be required for equivalent models.

in both the cohort parameters and $\kappa_t^{(1)}$ which will need to be allowed for in projection. However small they may be in the historical data, these cubic trends will eventually come to dominate the long term evolution of mortality rates, potentially yielding projected mortality rates which lack biological reasonableness due to apparent changes in trend.

Consequently, it may be prudent to avoid unidentifiable cubic (and higher) order polynomial trends in an APC mortality model. Such trends arise when we use more complicated models with higher-order polynomial age functions. It is therefore useful, when selecting such models, to have a larger "toolkit" of age functions for use in the models than simply extending existing models by using higher-order polynomial terms. Chapter 5 proposed such a toolkit, which allows for more complicated mortality models that do not suffer from excessive identifiability issues and can give biologically reasonable, well-identified projections of mortality rates, as shown in Chapters 6 and 8.

4.6 Conclusions

In Chapter 3, we saw how AP mortality models are not fully identified, and that in order to identify these models, most users impose additional arbitrary identifiability constraints on them when fitting the models to data. Some APC mortality models have extra identifiability constraints, caused by the collinearity between age, period and cohort, which are unlike anything found in similar AP models. These depend upon the form of the age functions in the model and so are specific to individual models. The identifiability issues involve deterministic trends which cannot be uniquely allocated between the age, period or cohort terms and so an arbitrary allocation must be made via additional arbitrary identifiability constraints. The nature of the unidentifiable trends present in specific models are summarised in Figure 4.1.

These unidentifiable deterministic trends have important consequences when we come to project the model. We must first determine the identifiability issues in the specific model we are using, in order to find which deterministic trends are unidentifiable. When this is done, we can specify suitable time series processes for the variation around these trends. Only by doing this can we ensure that our projected mortality rates are independent of the arbitrary identifiability constraints imposed when fitting the model.

By understanding these identifiability issues, however, we can build more complex mortality models, for instance, via the "general procedure" of Chapter 5, and be confident that they are founded on a secure knowledge of the underlying mathematical structure of APC mortality models. We are also able to use more sophisticated time series projection methods, as in Chapters 6 and 8, knowing that our projections are free from dependence on the arbitrary choices we made when fitting the model to data.

4.A Identifiability in APC models with non-parametric age functions

The matrix form of AP mortality models, given by Equation 3.3 in Chapter 3, can be extended to allow for cohort effects

$$H = \alpha \mathbf{1}_T^\top + \beta\kappa + \gamma \tag{4.36}$$

where γ is an $(X \times T)$ Toeplitz matrix, i.e., a matrix where the diagonal elements are constant. It is clear that the transformations in Equations 3.11 and 3.12 are still invariant transformations of Equation 4.36 and therefore the conclusions of Chapter 3 are still applicable in the wider context of APC mortality models. Indeed, the transformation in Equation 4.4 of the classic APC model is simply the transformation in Equation 3.12 applied to this specific model.

Generalising Equation 4.5 in this context for more complicated invariant transformations, we can see that the transformation

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{\gamma}\} = \{\alpha - c\mathbf{1}_X, \beta, \kappa, \gamma + c\mathbf{1}_X \mathbf{1}_T^\top\} \tag{4.37}$$

is common to all APC models of the form in Equation 4.36 (where $\mathbf{1}_X$ is a $(X \times 1)$ column vector of ones and similarly for $\mathbf{1}_T$). This transformation was also discussed (using alternative notation) in Section 4.4. This allows us to set the level of the cohort parameters - typically to be around zero to impose the demographic significance discussed in Chapter 2.

To generalise the transformation in Equation 4.6, if we can find a Toeplitz matrix Γ such that[35]

$$\Gamma = a\mathbf{1}_T^\top + \beta k \tag{4.38}$$

(with a an $(X \times 1)$ matrix and k an $(N \times T)$ matrix), we then have the transformation

$$\{\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{\gamma}\} = \{\alpha - a, \beta, \kappa - k, \gamma + \Gamma\} \tag{4.39}$$

In the case of the classic APC model, we have $\beta = \mathbf{1}_X$ and so can find a Toeplitz matrix $\Gamma = c(\mathbf{1}_X \boldsymbol{T}^\top - \boldsymbol{X}\mathbf{1}_T^\top)$ where $\boldsymbol{X}$ is the $(X \times 1)$ column vector $X_i = \{i - \bar{x}\}$ where i runs from 1 to X (and similarly for $\boldsymbol{T}$).

Theorem 4.3. *There are no invariant transformations of general APC mortality models with non-parametric age functions, i.e., no such A, k and Γ exist unless a specific shape for β is assumed in the model.*

Sketch of Proof Consider the general term $a\mathbf{1}_T^\top + \beta k$, which is analogous to the predictor structure of an AP mortality model. As we argue in Chapter 3, this has dimension $X + N(X + T) - N(N + 1)$, i.e., the X parameters in a, the NX parameters in β, and the NT in k reduced by the $N(N + 1)$ degrees of freedom in the transformations in Equations 3.11 and 3.12.

In contrast, in the general case, Γ has dimension $X + T - 1$, i.e., one degree of freedom for each diagonal. For Equation 4.38 to be true, these matrices must have the same dimension and therefore

$$\begin{aligned} X + N(X + T) - N(N + 1) &= X + T - 1 \\ N^2 + N(1 - X - T) + T - 1 &= 0 \end{aligned} \tag{4.40}$$

However, N, X and T are integers, set by the structure of the model and the range of the data, and therefore Equation 4.40 will not generally be true. Hence Equation 4.39 will not be an invariant transformation of a general APC mortality model with non-parametric age functions.

[35] We actually require the more general statement that $\Gamma = a\mathbf{1}_T^\top + bk$ with b a $(X \times N)$ matrix such that $\beta = bA$, i.e., the columns of b lie within the span of the columns of β. However, without loss of generality, we define $\tilde{k} = Ak$ to obtain the result shown.

The argument used in this proof relies on $a\mathbf{1}_T^\top + \beta k$ being of full rank and therefore breaks down if β is of lower dimension than the maximum possible. However, this is equivalent to imposing a parametric form on the age functions and accordingly, the line of reasoning above is not possible in the general case.

Therefore, general non-parametric APC mortality models do not possess any other invariant transformations apart from the ones in Equations 3.11, 3.12 and 4.37. They require only identifiability constraints which set the normalisation scheme of the age functions, impose orthogonality between the age and period functions (both using the transformation in 3.11), set the levels of the period functions $\kappa_t^{(i)}$ using Equation 3.12, and the level of the cohort parameters γ_{t-x} using Equation 4.37. □

For instance, we see that for the H1 model of Haberman and Renshaw (2009) and Hunt and Villegas (2015),

$$\eta_{x,t} = \alpha_x + \beta_x \kappa_t + \gamma_{t-x} \tag{4.41}$$

we cannot find an invariant transformation of the parameters similar to that in Equation 4.6. This is because of the lack of shape in either age or period in the $\beta_x \kappa_t$ term which can be used to decompose the cohort term. However, this model does possess an "approximate" identifiability constraint, which leaves the fitted mortality rates almost unchanged in the majority of cases. This is caused by κ_t often having a form that is close being parametric, which is discussed in detail in Hunt and Villegas (2015).

Some, especially demographers, have argued that all cohort effects are simply misspecified age/period effects and are best modelled as such.[36] Although this may be true in a strictly mathematical sense, a large number of age/period terms are required to replicate any general cohort term in the model. It is therefore more parsimonious to include a set of cohort parameters rather than multiple age/period terms. This, again, is similar to the argument in Wilmoth (1990), which states that it is plausible and parsimonious to include a single set of cohort parameters rather than an excessive number of age/period terms which achieve the same effect.

Some datasets may show little or no structure across years of birth, in which case the decision to include a cohort term becomes one decided on the basis of the demographic and statistical significance of the parameters for that dataset. Such a decision can be

[36] For instance, Cairns et al. (2011a) raised *"the possibility that cohort effects might be partially or completely replaced by well-chosen age and period effects"* and also see Murphy (2010)

made only after all significant age/period terms have been identified. We therefore recommend a procedure, such as the "general procedure" in Chapter 5, which only adds such a term when justified by the data.

4.B Models without a static age function

As we discuss in Chapter 2, a number of APC mortality models have been proposed which do not have an explicit static age function, α_x, the most prominent of which being the extensions of the CBD model in Cairns et al. (2009). If the model does not have an explicit static age function, the age functions in the model must be parametric and therefore known in advance of fitting the model to data. The structure of the APC model in this case is therefore

$$\eta_{x,t} = \sum_{i=1}^{N} f^{(i)}(x)\kappa_t^{(i)} + \gamma_{t-x}$$

The identifiability issues in such models can be considered in the same fashion as in Section 4.4. In particular, we noted in Section 4.4.2 that the invariant transformations of models with exponential or trigonometric age functions did not involve the static age function, and therefore are also applicable in models without one.

The invariant transformations of models with polynomial age functions, in contrast, did involve the static age function explicitly. The proof of Theorem 4.1 involves expanding a polynomial function of year of birth, $g(y)$, into polynomial terms in x and t and then combining these in the appropriate age/period terms. In particular, the term in this expansion with no t dependence was combined into the static age function. This is seen most clearly in the transformation in Equation 4.6, but also in the transformation in Equation 4.13 for the Plat model.

However, we can see that the lack of a static age function to absorb this term in the expansion of $g(y)$ is not an insurmountable problem as long as there is an age/period term with the appropriate age function. This means that if $g(y)$ is a polynomial of order M, we must have age functions in the model up to order M as well. This contrasts with models with a static age function, which only require age functions up to order $M-1$.

Theorem 4.4. *APC mortality models with no static age function and age functions spanning the polynomials to order M possess invariant transformations which adds a polynomial of order M to the cohort function.*

Sketch of Proof The proof is similar to that of Theorem 4.1. Take $g(y)$, a general polynomial of order M, and expand as a function of x and t. This can then be regrouped into an equivalent form that corresponds to the age/period terms in the model, in order to see how $g(y)$ can be absorbed into the age/period structure

$$\begin{aligned}
g(y) &= \sum_{n=0}^{M} a_n y^n \\
\Rightarrow g(t-x) &= \sum_{n=0}^{M} a_n (t-x)^n \\
&= \sum_{n=0}^{M} a_n \sum_{m=0}^{n} \binom{n}{m} t^m (-x)^{n-m} \\
&= \sum_{n=0}^{M} \sum_{l=0}^{n} a_n \binom{n}{l} t^{n-l} (-x)^l \\
&= \sum_{l=0}^{M} (-x)^l \sum_{n=l}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{l=0}^{M} (-1)^l f^{(l)}(x) \sum_{n=l}^{M} a_n \binom{n}{l} t^{n-l} \\
&= \sum_{l=0}^{M} f^{(l)}(x) k^{(l)}(t)
\end{aligned}$$

which is of the form of Equation 4.9 if the age functions in the model are of the form $f^{(j)}(x) = x^j$ of $j = 0, 1, \ldots M$. □

To see this in practice, consider model M6 of Cairns et al. (2009)

$$\eta_{x,t} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \gamma_{t-x} \tag{4.42}$$

and compare it with the reduced Plat model of Equation 4.12 in Section 4.4.1.1. For the reduced Plat model, we saw that the transformation in Equation 4.13 was invariant, and involved adding a quadratic function of year of birth to the cohort parameters, with adjustments to $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and the static age function α_x. For model M6, this transformation is not permitted, as there is no static age function to adjust in this model. Instead, the model only has the simpler linear invariant transformation

$$\{\hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\kappa_t^{(1)} - c(t - \bar{t}), \kappa_t^{(2)} - c, \gamma_y - c(y - \bar{y})\} \tag{4.43}$$

We can also see this using the analysis of Chapter 2, where it was shown that models without a static age function can be written as though they do have one of a specific, parametric form that has been combined with the other age/period terms in the model.

In the case of model M6, we see that this implies a static age function which is a linear function of age, which then could not be used to absorb a quadratic age term coming from the addition of a quadratic function of year of birth to the cohort parameters. Consequently there is a trade-off: models without a static age function have simpler identifiability issues than (otherwise similar) models possessing one, but are unable to provide a good fit to mortality data across the full age range, as discussed in Chapter 2.

4.C Maximal invariants

An alternative approach to using an arbitrary identification scheme was suggested by Kuang et al. (2008b,a) and Nielsen and Nielsen (2014) for the classic APC model. This is to change the parameterisation of the model to an equivalent form with reduced dimensionality which does not suffer from identifiability issues. The new parameters are known as "maximal invariant" parameters, since they are the set with the largest number of parameters (i.e., are "maximal"), and are injective[37] and give the same fitted mortality rates as the original model in Equation 4.1 (i.e., the reparameterisation is "invariant") . We can think of this as finding a parameterisation of the model which gives the same fit to data, but where every possible degree of freedom in the model is fully utilised in fitting the data.

Kuang et al. (2008b) and Nielsen and Nielsen (2014) proposed an approach to generating a maximally invariant parameterisation for the classic APC model based on finding the second differences of the age, period and cohort terms. These second differences do not change under the invariant transformations of the model and so have a meaning independent of the identifiability constraints. In this Appendix, we review this approach and discuss how it can be extended to deal with the identifiability issues in some of the more complex APC mortality models. However, we also find that it suffers from a number of limitations which make it unsuitable for many APC models and which can cause projections to be biologically unreasonable.

[37] A transformation is injective if different points in the domain get mapped to different points in the image of the transformation.

First, the age, period and cohort functions in the classic APC model are expanded as telescopic sums in terms of their second differences, i.e.,

$$\begin{aligned}
\alpha_x &= \alpha_X - \sum_{i=x+1}^{X} \Delta\alpha_i \\
&= \alpha_X - \sum_{i=x+1}^{X} \left(\Delta\alpha_X - \sum_{j=i+1}^{X} \Delta^2\alpha_j \right) \\
&= \alpha_X - (X-x)\Delta\alpha_X + \sum_{i=x+1}^{X} \sum_{j=i+1}^{X} \Delta^2\alpha_j \\
\kappa_t &= \kappa_1 + (t-1)\Delta\kappa_2 + \sum_{i=2}^{t} \sum_{j=3}^{t} \Delta^2\kappa_j \\
\gamma_y &= \gamma_{1-X} + (y-1+X)\Delta\gamma_{2-X} + \sum_{i=2-X}^{y} \sum_{j=3-X}^{y} \Delta^2\gamma_j
\end{aligned}$$

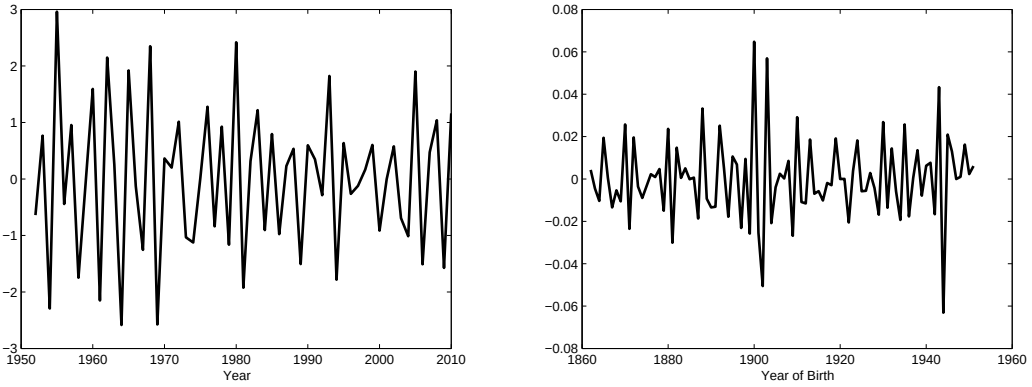
In the case of the age function, α_x, we work backwards from α_X due to the negative dependence of cohort on age. However, it is important to note that this expansion has not changed the number of parameters in the model, merely written them in a new form. This, of itself, will not solve the identifiability issues. However, Kuang et al. (2008b) and Nielsen and Nielsen (2014) then substituted the second difference expansions of the parameters into the classic APC model and group the deterministic terms together

$$\eta_{x,t} = a_0 + (X-x)a_1 + (t-1)b_1 + \sum_{i=x+1}^{X} \sum_{j=i+1}^{X} \Delta^2\alpha_j + \sum_{i=2}^{t} \sum_{j=3}^{i} \Delta^2\kappa_j + \sum_{i=2-X}^{t-x} \sum_{j=3-X}^{i} \Delta^2\gamma_j \tag{4.44}$$

where

$$\begin{aligned}
a_0 &= \alpha_X + \kappa_1 + \gamma_{1-X} \\
a_1 &= \Delta\gamma_{2-X} - \Delta\alpha_X \\
b_1 &= \Delta\kappa_2 + \Delta\gamma_{2-X}
\end{aligned}$$

In Kuang et al. (2008b) and Nielsen and Nielsen (2014), these new parameters were introduced by considering three points of the fitted mortality surface. The most important point about the procedure is that it replaces six parameters in the original parameterisation with only three in the maximally invariant parameterisation. The maximally invariant parameterisation therefore contains $3+(X-2)+(T-2)+(T+X-3) = 2X+2T-4$ free parameters. This compares with $2X+2T-1$ parameters and the three additional identifiability constraints required by the three invariant transformations - Equations 4.4, 4.5 and 4.6 - for the original parameterisation of the classic APC model. Hence the



(A) $\Delta^2\kappa_t$ (B) $\Delta^2\gamma_y$

FIGURE 4.4: Second differences from the classic APC model

maximally invariant parameterisation gives the same fitted mortality rates with the same number of effective parameters but without the over-parameterisation and consequent need for identifiability constraints in the original formulation of the model.

However, by doing this, we have lost much of the demographic significance associated with the original parameters in the classic APC model. For example, whilst α_x in the original parameterisation of the classic APC model relates to an age effect specific to age x, $\Delta^2\alpha_x$ relates to the curvature of the mortality curve in the age dimension at age x and will impact mortality rates at all ages below x. It is therefore harder to explain its demographic significance to other model users or develop an intuition about what values are reasonable in order to check the validity of the model. Although demographic significance is subjective, it is still not desirable to lose it if it can be avoided. This may restrict the usefulness of the maximally invariant approach.

In order to project the model into the future, we need to analyse the $\Delta^2\kappa_t$ and $\Delta^2\gamma_y$ parameters as time series. These are shown in Figure 4.4 for the same dataset as used in Section 4.5.2. As can be seen,[38] these parameters appear to be stationary and so it is natural to project them using an ARMA process.

If we were to "integrate up" the double differences to recover our original κ_t and γ_y parameters, these would both be $I(2)$ processes. This conflicts with the demographic significance for the cohort parameters discussed in Chapter 2. $I(2)$ processes are also not

[38] We have removed the large outlier cohort effects for years of birth 1918/19 using indicator variables, as they are believed to be data artefacts resulting from the surge of births due to the demobilisation of soldiers after the First World War, based on similar reasons as those presented in Richards (2008) and Cairns et al. (2014).

likely to be biologically reasonable, as the uncertainty in projected mortality rates would grow very quickly. This would have important ramifications if the model is projected.

The maximal invariant approach also works with some other APC mortality models. For instance, consider the reduced Plat model of Equation 4.12. This model has $X + 2T + (X + T - 1) = 2X + 3T - 1$ parameters and, as discussed in Section 4.4.1.1, we know that it requires five identifiability constraints to fully identify (two for the level of the period functions and one each for the level, linear trend and quadratic trend in the cohort parameters).

In order to find a maximally invariant parameterisation, we follow the same logic as in Kuang et al. (2008b) and consider the telescopic sums of the parameters. However, as α_x, $\kappa_t^{(1)}$ and γ_y all possess unidentifiable quadratic trends, we need to consider the third differences of these parameters, but only consider the second differences of $\kappa_t^{(2)}$, since it only has unidentifiable linear trends

$$\begin{aligned}
\alpha_x &= \alpha_X - (X-x)\Delta\alpha_X + \frac{1}{2}(X-x)(X-1-x)\Delta^2\alpha_x - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k \\
\kappa_t^{(1)} &= \kappa_1^{(1)} + (t-1)\Delta\kappa_2^{(1)} + \frac{1}{2}(t-1)(t-2)\Delta^2\kappa_3^{(1)} + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)} \\
\kappa_t^{(2)} &= \kappa_1^{(2)} + (t-1)\Delta\kappa_2^{(2)} + \sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)} \\
\gamma_y &= \gamma_{1-X} + (y-1+X)\Delta\gamma_{2-X} + \frac{1}{2}(y-1+X)(y-2+X)\Delta^2\gamma_{3-X} \\
&\quad + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k
\end{aligned}$$

Combining these in Equation 4.12 and grouping the deterministic terms of the same type reduces the dimension of the parameter set in the same manner as for the classic APC model. Therefore, we find the maximally invariant form of the reduced Plat model

$$\begin{aligned}
\eta_{x,t} &= a_0 + (x-\bar{x})a_1 + (x-\bar{x})^2a_2 + (t-\bar{t})b_1 + (t-\bar{t})^2b_2 + (x-\bar{x})(t-\bar{t})c_1 \\
&\quad - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)} + (x-\bar{x})\sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)} \\
&\quad + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k
\end{aligned} \tag{4.45}$$

The final step to prove that this is a maximally invariant parameterisation would be to check that each of the parameters can be estimated uniquely from the data. Alternatively and more easily, we can see that it is maximally invariant from a dimensional argument, since the parameterisation has $6+(X-3)+(T-3)+(T-2)+(X+T-4) = 2X+3T-6$ free parameters, which is the same as the number of parameters in the original reduced Plat model less the number of identifiability constraints imposed. Therefore, the freely varying parameter space has the same dimension as the model space and gives the same fitted mortality rates as the original model, and so the parameters represent maximal invariants. Because of this, the revised model does not possess any identification issues.

As in the case of the classic APC model, moving to a maximally invariant form for the model means losing the demographic significance of the parameters. The maximally invariant form of the reduced Plat model is highly unintuitive compared with the original parameterisation, and it would be difficult to communicate the impact of the various parameters to anyone not intimately familiar with the maximally invariant approach. As discussed in Chapter 2, since demographic significance is a major reason for choosing a model with parametric, as opposed to non-parametric age functions, this is highly undesirable. Also, and again similar to the classic APC model, the use of third differences for $\kappa_t^{(1)}$ and γ_y leads naturally to using $I(3)$ processes when we project the model, which are unlikely to give biologically reasonable projections.

Further, the maximal invariant approach does not work with all APC mortality models. If we follow the same logic to try to find the maximally invariant parameterisation for the full Plat model in Equation 4.11 we obtain

$$\begin{aligned}\eta_{x,t} = {} & a_0 + (x-\bar{x})a_1 + (x-\bar{x})^2 a_2 + (t-\bar{t})b_1 + (t-\bar{t})^2 b_2 + (x-\bar{x})(t-\bar{t})c_1 \\ & - \sum_{i=x+1}^{X}\sum_{j=i+1}^{X}\sum_{k=j+1}^{X}\Delta^3\alpha_k + \sum_{i=2}^{t}\sum_{j=3}^{t}\sum_{k=4}^{t}\Delta^3\kappa_k^{(1)} + (x-\bar{x})\sum_{i=2}^{t}\sum_{j=3}^{t}\Delta^2\kappa_j^{(2)} \\ & + (x-\bar{x})^+\kappa_1^{(3)} + (x-\bar{x})^+\sum_{i=2}^{t}\Delta\kappa_i^{(3)} + \sum_{i=2-X}^{y}\sum_{j=3-X}^{y}\sum_{k=4-X}^{y}\Delta^3\gamma_k \end{aligned} \tag{4.46}$$

We know, from Section 4.4.1.1, that the Plat model has $X+3T+(X+T-1) = 2X+4T-1$ parameters and requires six identifiability constraints (three on the levels of the period functions and one each for the level, linear trend and quadratic trend in the cohort parameters). However, the maximally invariant parameterisation in Equation 4.46 has $7+(X-3)+(T-3)+(T-2)+(T-1)+(X+T-4) = 2X+4T-6$ free parameters, i.e., one too many. This is because the $(x-\bar{x})^+\kappa_1^{(3)}$ term cannot be combined with the expanded form of α_x, since it is not a polynomial. Consequently, there is no dimensional

reduction with respect to this age/period term.

Because of this, we will still require an additional identifiability constraint to fit the model in Equation 4.46 to data. However, it is no longer clear what this should be or what the underlying invariant transformation of the parameters is. The maximally invariant approach has therefore not solved the identifiability issues for this model, but has made making an arbitrary identification considerably more difficult.

This will be true for any age/period term which does not have a polynomial age function. As discussed in Section 4.4.3, such terms do not generate any additional identifiability issues beyond the unidentifiable level of the period function, as discussed in Chapter 3. It therefore may be possible to deal with this using an approach similar to that proposed for the model of Lee and Carter (1992) in Nielsen and Nielsen (2014) and discussed in the Appendix of Chapter 3. However, as these two techniques for obtaining maximally invariant parameterisations are fundamentally different, it is unclear how to combine them in models which mix polynomial and non-polynomial age functions, such as the Plat model.

In summary, the maximally invariant approach proposed in Kuang et al. (2008b) and Nielsen and Nielsen (2014) for the classic APC model can be generalised, but only to models with purely polynomial age functions. For models with other forms for the age functions (or which mix polynomial and non-polynomial age functions), the maximally invariant approach, at best, offers a partial solution. However, in using such an approach, we lose our desired demographic significance regarding the parameters in the model and are likely to obtain projected mortality rates which are not biologically reasonable, so this approach is not, in general, recommended.

Chapter 5

A General Procedure for Constructing Mortality Models

5.1 Introduction

In recent years, there has been an explosion in the number of new mortality models that have been proposed. This has been triggered, in part, by the greater focus placed on longevity risk by demographers, actuaries and governments. It has also been prompted by the failure of existing models to identify adequately the full extent of the complexities involved in the evolution of mortality rates over time.

Yet these new models often involve ad hoc extensions to existing models, which have questionable demographic significance.[1] Despite having more terms than the older models, they still fail to capture a lot of the information present in the data, such as the level of lifespan inequality in the population. They also have difficulties providing realistic forecasts of specific mortality rates. Lacking a formal procedure for interrogating the data to establish what structure remains to be explained, modellers too often add new terms based on theoretical models of mortality or on assumptions regarding the shape of the mortality curve rather than evidence. This is especially dangerous in models with cohort parameters intended to capture generational effects. The result of any mis-specification in these extra age/period terms can result in structure being wrongly attributed to the cohort effect. This is then projected incorrectly, moving up the age range with the passage to time, with the result that implausible forecasts are generated

[1] Demographic significance is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

at higher ages.

In view of this, we feel that the time has come to take a fresh look at mortality model construction. But, rather than propose yet another new model, what we do in this study is outline and implement a "general procedure" (GP) for building a mortality model from scratch, driven by a forensic examination of the data. Through an iterative process, the GP identifies every significant demographic feature in the data in a sequence, beginning with the most important. For each demographic feature, we need to apply expert judgement to choose a particular parametric form to represent it. To do this, we need a "toolkit" of suitable functions.

By following the GP, it is possible to construct mortality models with sufficient terms to capture accurately all the significant information present in the age, period and cohort dimensions of the data. In particular, the GP prevents structure in the data which is genuinely associated with an age/period effect being wrongly allocated to a cohort effect. The procedure is general in the sense that it can be applied to any dataset to give a fully specified model tailored to the features of the population under consideration. Most significantly, the GP provides evidence for the addition of each term to an existing model; it allows each new term to be associated with a specific demographic and biological process driving the evolution of mortality rates.

Section 5.2 presents a summary of the structure of the class of mortality models we are considering and sets out the desirable properties that we believe a good mortality model should possess. The general procedure is discussed in Section 5.3. In Section 5.4, we apply the GP to data for men in the UK and describe how the steps in Section 5.3 operate in practice. In Section 5.5, we assess the goodness of fit of this model and check whether there is any remaining structure present in the fitted residuals. Section 5.6 compares the GP with the Lee-Carter model and with a procedure based on principal component analysis as an alternative method of constructing mortality models with multiple age/period terms. Finally, Section 5.7 concludes with an assessment of how the final model found measures up against our set of desirable properties from Section 5.2 as well as its advantages and disadvantages.

5.2 The structural form of mortality models

The majority of existing mortality models proposed in the actuarial literature fall into an age/period/cohort framework. This transforms the observed mortality rates and then fits a series of terms to account for the interactions between the age, x, the year of observation, t, and the year of birth, $y = t - x$, for the population within each cell of data. Mathematically, this can be written as:[2]

$$\eta\left(\mathbb{E}\left(\frac{D_{x,t}}{E_{x,t}}\right)\right) = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x;\theta^{(i)})\kappa_t^{(i)} + \gamma_{t-x} \tag{5.1}$$

This equation has the following components:

- a link function η to transform the observed data into a form suitable for modelling. The raw data usually consists of death counts $D_{x,t}$ and exposures to risk $E_{x,t}$ at ages x and for years t;
- a static age function α_x to capture the general shape of the mortality curve that does not change with time;
- N age/period terms $f^{(i)}(x;\theta^{(i)})\kappa_t^{(i)}$, consisting of companion pairs of period terms $\kappa_t^{(i)}$ (or "trends") which give the evolution of mortality rates through time and age functions $f^{(i)}(x;\theta^{(i)})$ which determine which segments of the age range these trends affect; and
- cohort parameters γ_{t-x} which determine the lifelong effects that are specific to different generations as discussed in Willets (2004), denoted by their year of birth;

Many mortality models proposed to date can be written in this form. These include the Lee-Carter (LC) model proposed in Lee and Carter (1992) and extensions of this, such as those of Renshaw and Haberman (2003b) and Yang et al. (2010). It also includes the Cairns-Blake-Dowd family of mortality models (in Cairns et al. (2006a) and Cairns et al. (2009)), the classic age/period/cohort model of Hobcraft et al. (1982) and developments of these models such as the models proposed by Plat (2009a) and O'Hare and Li (2012a). In addition, it includes various other mortality models not contained within these families such as the ones proposed in Wilmoth (1990) and Aro and Pennanen (2011). The models of the rate of mortality change proposed in Haberman and Renshaw (2012, 2013)

[2] This structural form and demographic significance of the terms in it are discussed in depth in Chapter 2.

and Mitchell et al. (2013) also fall within this structure for suitable choice of the link function $\eta_{x,t}$. These models and the relationships between them are discussed in greater depth in Chapter 2. Examples of models which fall outside this framework include those with a constant, Makeham term, the extension to the LC model proposed in Renshaw and Haberman (2006) (due to the presence of the $\beta_x^{(0)}$ term modifying the cohort parameters) and the P-splines models of Currie et al. (2004).

A good mortality model should satisfy the following "desirability criteria":

1. provide an adequate fit to the data, with sufficient terms to capture all the significant structure in the data;
2. be biologically reasonable;[3] and have terms which have demographic significance in the sense that they are explainable in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates at specific ages
3. be parsimonious, with the smallest number of terms needed to capture this structure, and with each term using as few parameters as possible;
4. be robust, in that parameter uncertainty should be low and small changes in the data should not result in significant changes in the estimates of the parameters and in our interpretation of them;
5. span the full age range, with sufficient terms to model the complex shape of and dynamics observed in mortality rates at younger ages; and
6. include cohort effects if justified by the data and allow for these to be clearly distinguished from age/period effects to allow plausible projections of the model.

The GP has been designed with these criteria (and the trade-offs between them) in mind. Most specifically, the GP chooses parametric age functions,[4] $f^{(i)}(x;\theta^{(i)})$, which take a specific functional form and are parameterised by a small number of variables $\theta^{(i)}$, over more general non-parametric age functions,[5] $\beta_x^{(i)}$, due to their parsimony and because we can use our judgement to assign demographic significance to the term in question. The advantages and disadvantages of using parametric age functions are discussed in greater depth in Chapter 2. However, a key feature of the GP is to use the information discovered from first using a non-parametric age function to provide guidance on the shape of that demographic feature. This will improve the goodness of fit for each term and avoid the need to make a priori assumptions regarding which age functions to use.

[3] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

[4] Defined in Chapter 2 as one taking a specific functional form that is defined by an algebraic formula.

[5] Defined in Chapter 2 as one fitted without imposing any a priori structure across ages.

5.3 A general procedure for constructing mortality models

The general procedure consists of the following steps:

1. Start with a static age function α_x to capture the time-independent shape of the mortality curve across ages in the data set under consideration;
2. Add a companion pair of non-parametric age and period functions $\beta_x\kappa_t$ to find the most significant age/period effect not captured by the model so far, where the age term β_x is free to take the shape that maximises the fit to the data;
3. Observe the shape of the estimated age term β_x across ages and how κ_t has evolved through time;
4. Check that the addition of the new pair of terms improves the overall goodness of fit to the data;
5. Use judgement to select a specific smooth functional form $f(x;\theta)$ to replace the non-parametric age term β_x where the function is defined by a small number of free parameters θ;
6. Check whether the fitted model with this specific functional form
   (a) produces a similar evolution over time as the non-parametric term by comparing the fitted κ_t's for the two cases and
   (b) achieves comparable improvements in the goodness of fit as the non-parametric term.
7. Check whether the addition of the new companion pair of terms has significantly changed the shape of previously selected terms, in which case we might need to change and re-estimate the earlier terms;
8. Repeat steps 2 to 7 until we are satisfied that the model captures all significant age and period structure in the data;
9. Add a cohort term γ_{t-x} to capture any year of birth effects;
10. Test the final model for goodness of fit and robustness, and the residuals for the properties of normality and independence, thereby confirming that there is no significant unexplained demographic structure remaining in the data;
11. Compare the final model to alternative models estimated using the same data set.

After each modification of the model structure (e.g., replacing a non-parametric age function, β_x, with a parametric alternative, $f(x)$, or the addition of the cohort term), all the terms are re-estimated by fitting the model to historical data.[6] This ensures that all of the parameters are estimated on the basis of maximising the fit to data and that there is no explicit hierarchy within the model structure. Figure 5.1 shows a flow chart of the GP summarising these steps.

The GP is a data-driven procedure, with terms being selected based on their ability to capture features of the observed mortality rates. At high level, it is a specific-to-general model building procedure (as defined in Campos et al. (2005)) as it begins with a simple model and sequentially adds terms in order to build a model that fully reflects the features contained in the dataset under investigation. This approach is unavoidable, since to begin with a fully general mortality model, as required by the general-to-specific methodology, would contain such a large number of terms that it would be impossible to fit it to data and difficult to simplify. However, at the "micro" level, each age/period companion pair is added in a general-to-specific fashion - the most general form of the function is added to the model and then simplified into a specific, parametric form, whilst seeking to retain its explanatory power. Thus, we believe that the GP benefits from both model-building frameworks.

The GP selects the functional form of the age/period terms in two stages. First, it allows each age/period term within the data to be identified by a non-parametric age function without requiring any a priori assumptions to be made by the modeller. Second, it allows the shape of these non-parametric age functions to guide the choice of parametric function that is selected from the toolkit to match as closely as possible the explanatory power of the former, whilst benefiting from parsimony in terms of the number of parameters to be estimated. However, judgement is required in the selection of the parametric function, although that the GP provides evidence to justify the decision made.

Appendix 5.A gives details of the "toolkit" of parametric age functions needed to implement the GP; it also gives a general algorithm for estimating the free parameters in them. However, a toolkit is never complete and so we do not offer this as an exhaustive list of functions - only as those we have considered so far. Two highly desirable features for a function to be included in the toolkit are a small number of free parameters (in our experience, more than two free parameters leads to unstable estimates) and the ability

[6] The only exception to this is when an exploratory $\beta_x\kappa_t$ term is added to the model, since these models are often very unstable due to over-parametrisation.

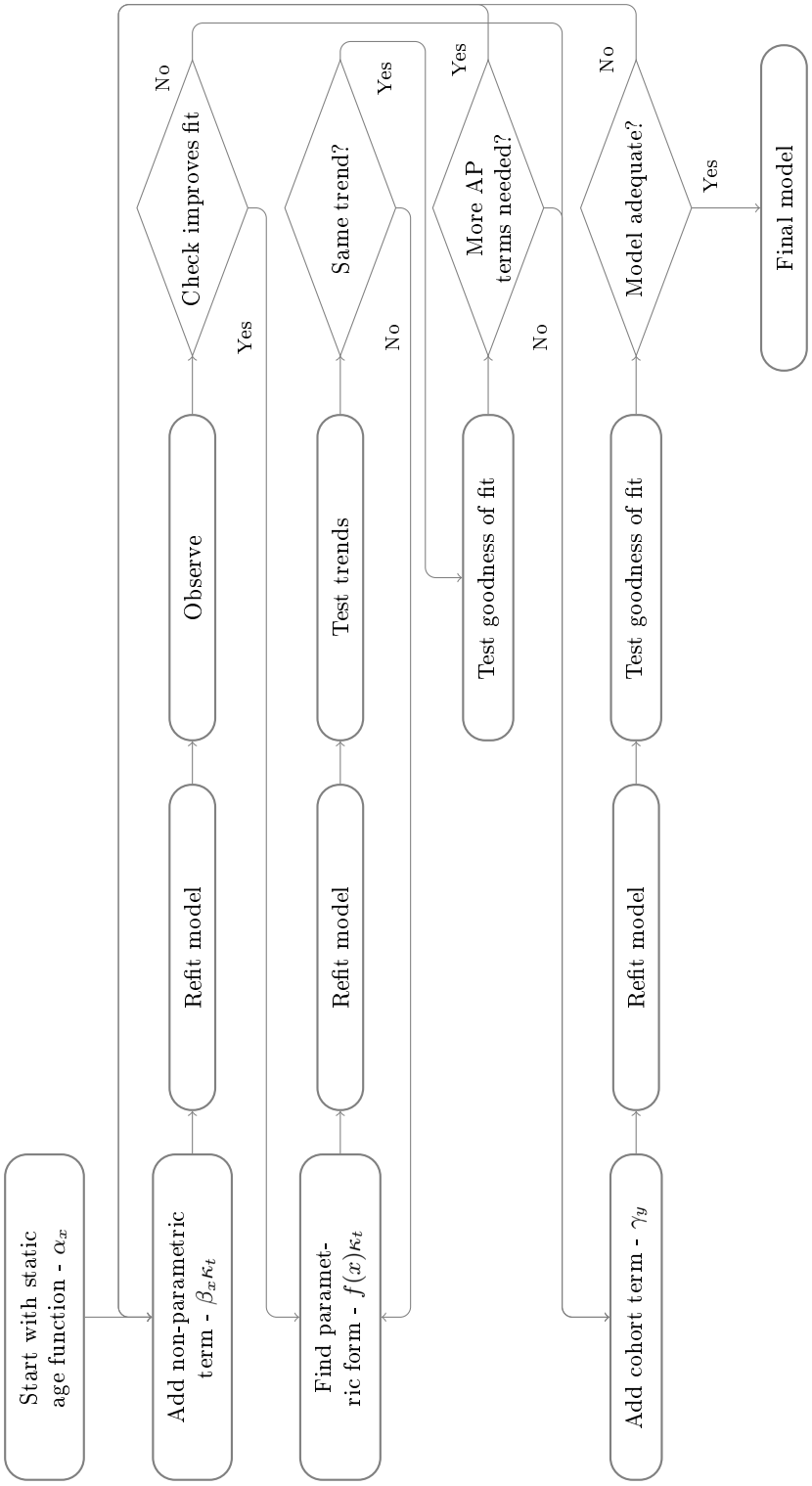


FIGURE 5.1: Flow chart of the general procedure

to adjust the location of the function in the age range.

At each stage of the GP, we need to assess whether the resulting model is in accordance with our desirability criteria. First, we will need to test whether an additional age function improves the fit of the model to data. It is well known that a measure such as the log-likelihood will always show an improvement in the fit of a series of nested models to the data due to the increased number of free parameters. In order to achieve our desire for a parsimonious model, it is therefore necessary to penalise the number of free parameters used by considering a measure such as the Bayes Information Criterion (BIC).[7] The log-likelihood is still useful, however, when adding an additional non-parametric term as the change in this measure represents the maximum possible improvement in the fit from the addition of a single new term. We can therefore use this maximum possible improvement as the benchmark for measuring the success of the specific parametric form being trialled: a parametric age function which produces 80-90% of the same improvement in log-likelihood can be regarded as highly desirable.

Second, we need to compare whether the structure identified by a non-parametric age function is the same as that found when a specific parametric function is introduced. Plots of the two are useful for revealing the general pattern of mortality change and identifying features such as trend changes and outliers that the two series have in common.

Finally, we will need to test the residuals from the data. As discussed in Pitacco et al. (2009), under a Poisson model for deaths (such as the one we use), the standardised deviance residuals $r_{x,t}$ are given by

$$r_{x,t} = sign(d_{x,t} - \hat{d}_{x,t})\sqrt{\frac{2W_{x,t}}{\phi}\left(d_{x,t}\ln\left(\frac{d_{x,t}}{\hat{d}_{x,t}}\right) - (d_{x,t} - \hat{d}_{x,t})\right)}$$

with actual death count $d_{x,t}$, fitted death count $\hat{d}_{x,t} = E^c_{x,t}\mu_{x,t}$, and ϕ the scale parameter given by the total fitted deviance divided by the number of degrees of freedom[8] of the model. This assumes that the residuals have constant variance across age and time. For large expected death counts, these should be approximately standard normal variables, so we can test the residuals for normality using the Jarque-Bera test of the skewness and kurtosis to check this. The residuals should also be independent and show no obvious structure across ages, periods and cohorts. To look for structure within the residuals, we

[7] Defined as max(Log-likelihood) $- 0.5 \times$ No. free parameters $\times \ln$(No. data points).
[8] Number of data points less number of free parameters.

plot heat maps and visually inspect for obvious vertical, horizontal or diagonal banding patterns. This would indicate the presence of further age, period or cohort effects. We also calculate the correlations of the residuals with their neighbours in the age and period directions, and test these correlations against the assumption of independence.

To exit the cycle of adding new age/period terms, we need a stopping rule in the GP to determine when there are no further demographically significant age/period terms left unidentified in the data. Such a stopping rule will inevitably be subjective. This means that the GP is not a "black-box" algorithm; it requires the active engagement and exercise of judgement by the modeller at each stage of the model building process.

Finally, we add the cohort parameters as the last step in the GP. The reason for this reflects a preference for a model where the majority of the temporal dependence in the data is allocated to the age/period terms. The reasons for this preference are discussed in detail in Chapter 2, but in our experience, the pattern of fitted cohort parameters produced by some models does not seem to have any demographic significance and may be caused by the model trying to compensate for inadequate age/period terms. We therefore seek to avoid this in the GP.

5.4 Application of procedure to male UK data

To illustrate the GP, we apply it to data for men in the UK from 1950 to 2009 covering ages 0 to 100 (ungrouped) downloaded from the Human Mortality Database (Human Mortality Database (2014)). We restrict the data to the period since the Second World War as it is free from major conflicts and abrupt social upheaval. Since the Human Mortality Database provides central exposures to risk for each age and year, we assume that the death counts are Poisson random variables and therefore use a log-link function for $\eta_{x,t}$ as it is the canonical link function for the Poisson distribution, as discussed in Chapter 2. We fit the model at each stage using Poisson maximum likelihood estimation using the algorithms described in Appendix 5.A.

5.4.1 Stage 0 - Static age function

The static age function produced by fitting $\ln(\mu_{x,t}) = \alpha_x$ constitutes the first step in the GP. The fitted values of α_x (not shown) show the usual pattern of mortality across the

full age range: with high mortality rates at age zero due to infant mortality, the log-linear pattern of mortality increases at high ages (from 50 to 90) and the increased rates of mortality due to the accident hump between ages 15 and 25. Whilst the age function is refitted at each stage of the GP, this shape does not change significantly throughout the different stages of the model building process.

5.4.2 Stage 1 - First age/period term

The next step is to add the first non-parametric age/period term to the static model to arrive at $\ln(\mu_{x,t}) = \alpha_x + \beta_x\kappa_t^{(1)}$, which has the form of the LC model. This gives the familiar β_x and $\kappa_t^{(1)}$ terms shown in Figure 5.2.

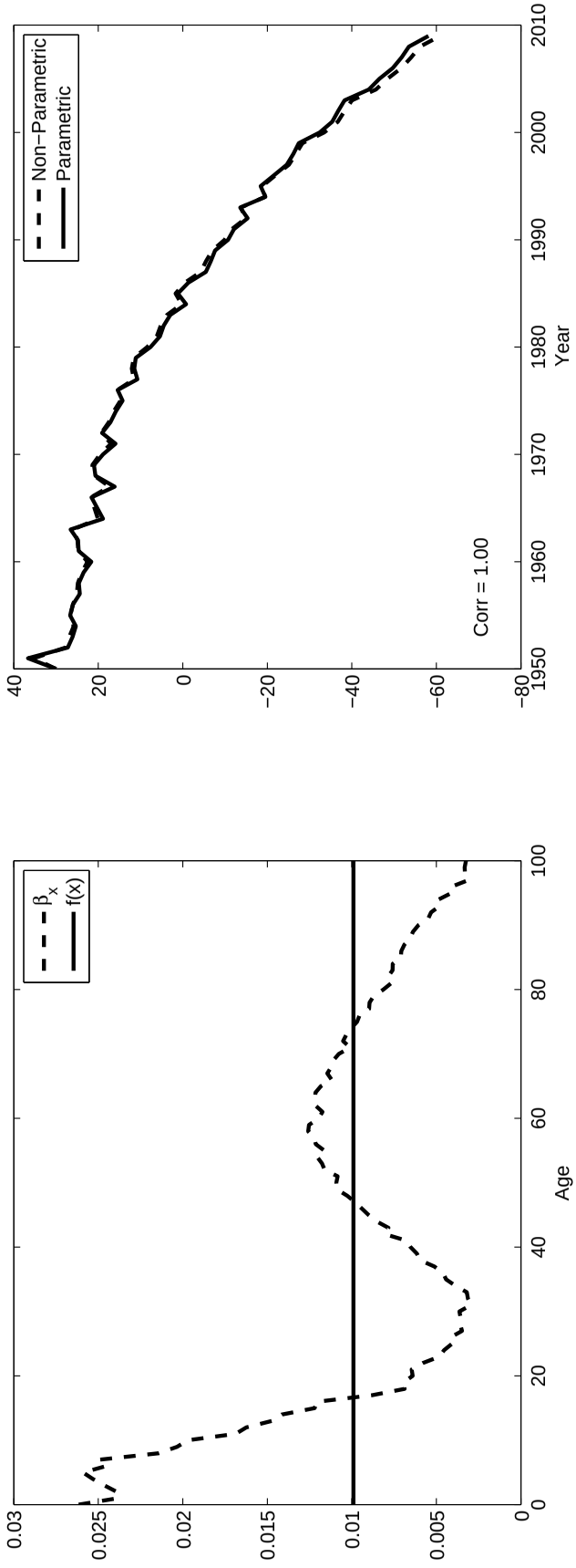
In order to fully identify the model, we impose

$$\sum_x |\beta_x| = 1 \tag{5.2}$$

$$\sum_t \kappa_t^{(1)} = 0 \tag{5.3}$$

and adopt these identifiability constraints for all subsequent age/period terms in the model for consistency. For parametric age functions, imposing Equation 5.2 involves rescaling the age function by either a constant or with a function of the free parameters, $\theta^{(i)}$ (i.e., ensuring that the age function is "self-normalising"). This is discussed further in Appendix 5.A and Chapter 3.

In the interests of parsimony and demographic significance, we believe that it is highly desirable to find a simpler parametric form than the age function of the LC model to capture the impact of the dominant trend within the data - ideally the simplest age function that will capture the same trend. This parametric form should be continuous to avoid any issues with the smoothness of projected mortality rates. As the fitted β_x age function is positive across the whole age range, it might be felt to represent a general improvement in mortality rates across all ages. Appealing to this demographic significance, we therefore try the simplest possible age function - a constant. As Figure 5.2 shows, this simple age function effectively captures the same trend as the non-parametric β_x function with 100 fewer parameters, and achieves approximately 92% of the same improvement in log-likelihood. We are therefore satisfied that there is no need to use a more complex and less parsimonious age function, although we would expect that much of the age structure present in the fitted β_x will need to be captured by subsequent age/period terms.



(A) Age functions

(B) Period functions

FIGURE 5.2: Age and period functions for Stage 1 of the general procedure

Figures 5.2a and 5.2b shows the age and period functions generated by Stage 1 of the GP. We can see that the population has experienced sustained improvements in mortality which have accelerated slightly in recent years. The model also detects the increased mortality in 1951 owing to the influenza epidemic in that year which affected much of England.

So far, so good, but a plot of the residuals - not shown here - indicates that additional terms are necessary to fully capture all the structure within the data.

5.4.3 Stage 2 - Second age/period term

In order to find the next most significant age/period effect within the data, we now add another non-parametric age/period term to the model to arrive at

$$\ln(\mu_{x,t}) = \alpha_x + f^{(1)}(x)\kappa_t^{(1)} + \beta_x\kappa_t^{(2)} \tag{5.4}$$

The fitted model gives the values of β_x and $\kappa_t^{(2)}$ shown in Figure 5.3. It is not a trivial task to select an appropriate parametric age function from the shape of β_x and this is where judgement becomes important. By inspection, the non-parametric age function appears to have two components - an upward-sloping linear trend across the entire age range and a large "hump" superimposed on the age range 10 to 50. Since we can assign different demographic significance to each of these features, it is appropriate that we separate them into two different age/period terms in the fully specified model. However, these trends will probably be highly correlated which is why the non-parametric function has combined them.

We choose to fit a straight line as our choice of $f^{(2)}(x)$ as it is a simpler potential function than one with a hump shape; indeed it is the simplest possible function after a constant. In our experience, a straight line is often the second choice of age function that arises naturally when applying the GP, especially for data restricted to higher ages. This lends support for the use of the Cairns-Blake-Dowd class of models. A straight line can be interpreted as determining changes in the slope parameter in a Gompertz model of mortality for models with a logarithmic link function. This is related to the "rectangularisation" of the mortality curve, as a greater proportion of deaths at high age occur around the median age of death. We also note that $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are negatively correlated, consistent with the Strehler-Mildvan law of mortality discussed in Finkelstein

(2012).

5.4.4 Stage 3 - Third age/period term

Our discussion of the choice of an appropriate age function at Stage 2 should give us a strong idea as to the appropriate shape of the age function for Stage 3. The GP gives us the evidence to support or reject our conjecture by first extending the model with a new non-parametric age/period term

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{2} f^{(i)}(x)\kappa_t^{(i)} + \beta_x \kappa_t^{(3)} \tag{5.5}$$

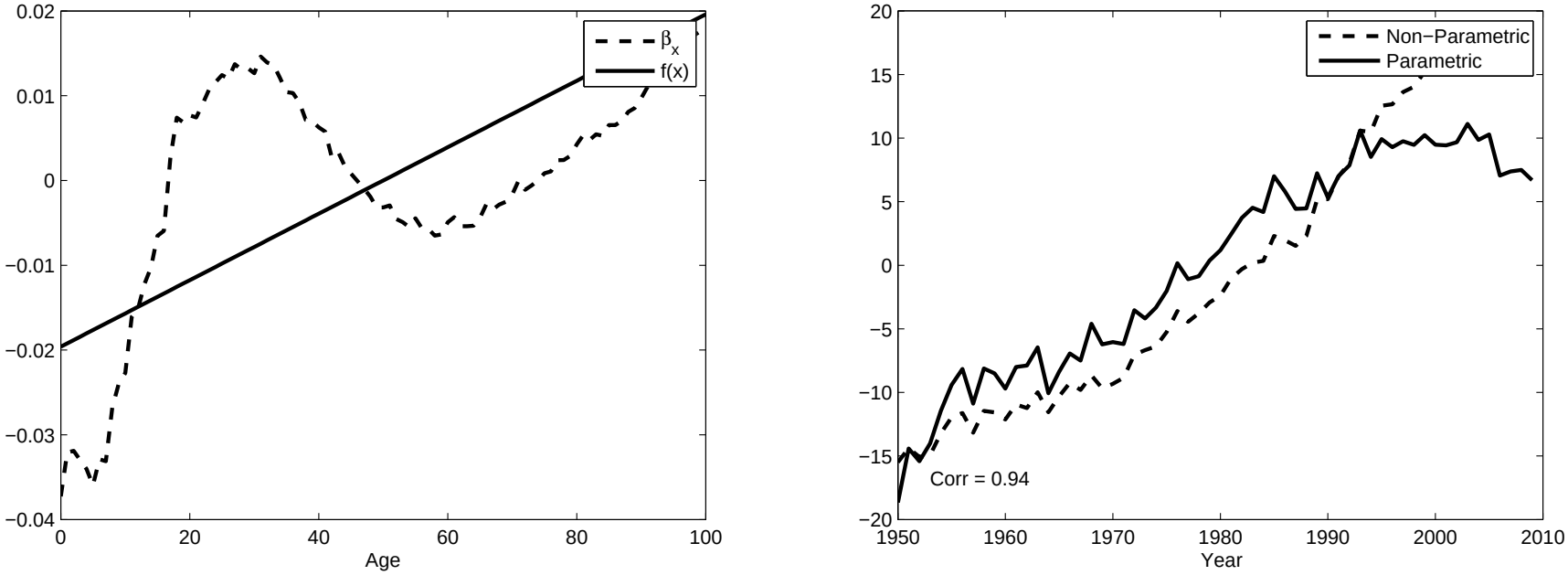
The fitted non-parametric model gives the values of β_x and $\kappa_t^{(3)}$ shown in Figure 5.4. This confirms that a suitable choice for $f^{(3)}(x)$ could indeed be some form of hump-shaped function centred around age 25 and so we experiment with

$$f^{(3)}(x;\hat{x},\sigma) \propto \frac{1}{\sigma}\exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right) \tag{5.6}$$

This function has two free parameters, $\hat{x}$ and σ which, by analogy with the normal distribution, govern the location of the hump and its width. These are estimated using Poisson maximum likelihood estimation. We choose the starting values for these parameters by observing the pattern of the β_x function, before applying our optimisation algorithm. The final, fitted values should not be overly sensitive to the initial choice. If they are, this indicates that the choice of age function may be inappropriate and will cause problems with the model when additional terms are added.

The final fitted $f^{(3)}(x;\hat{x},\sigma)$ and $\kappa_t^{(3)}$ functions are shown in Figure 5.4. When adding a new term to the model, we need to check that it does not significantly alter the demographic interpretation of the previous terms. Plots of the first two terms - not shown here - indicate that they have not changed significantly due to the presence of the third term.

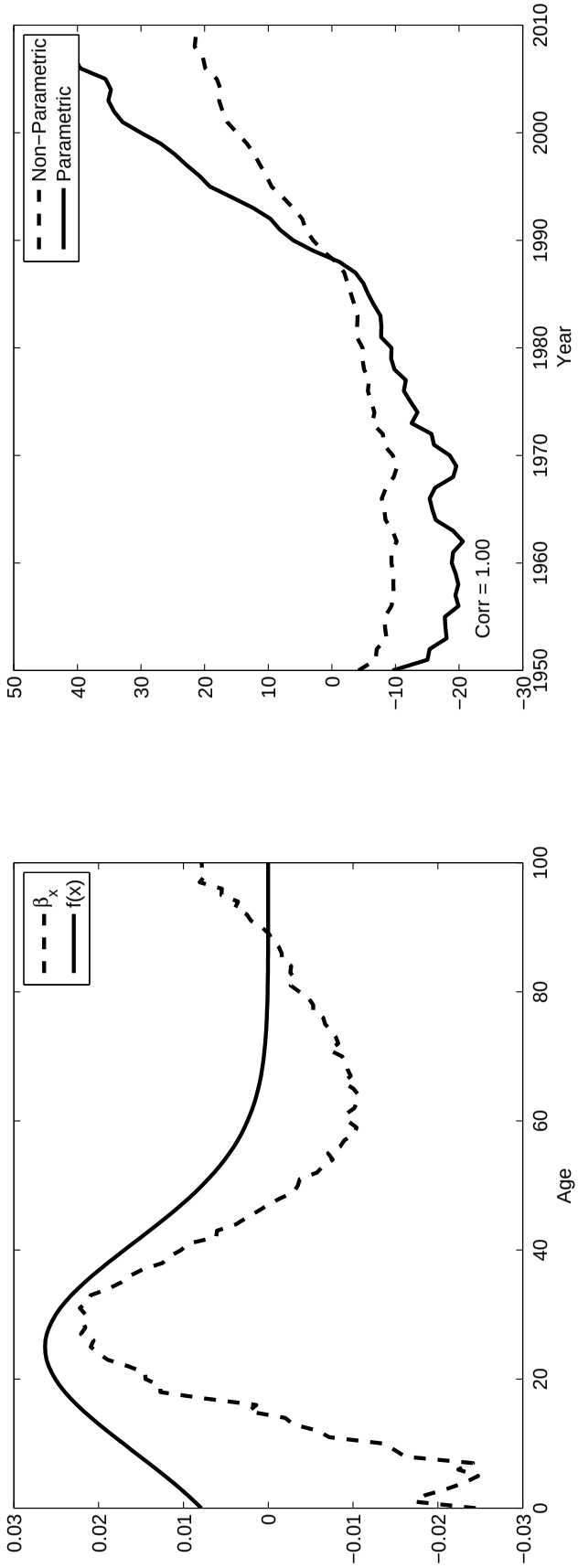
Visual inspection of the heat map of residuals in Figure 5.5 shows us that a) there appear to be additional age/period effects in the data, most obviously centred on age 0 and age 18 and b) there is a clear need for a cohort effect in the model as shown by the prominent diagonal lines on the heat map indicating features which follow individual years of birth as they age. The evidence gleaned from the heat map plot is useful when



(A) Age functions

(B) Period functions

FIGURE 5.3: Age and period functions for Stage 2 of the general procedure



(A) Age functions

(B) Period functions

FIGURE 5.4: Age and period functions for Stage 3 of the general procedure

deciding on subsequent terms, especially when trying to determine if the shape shown by an exploratory $\beta_x\kappa_t$ function is trying to approximate for a cohort effect - something we believe is essential to avoid.

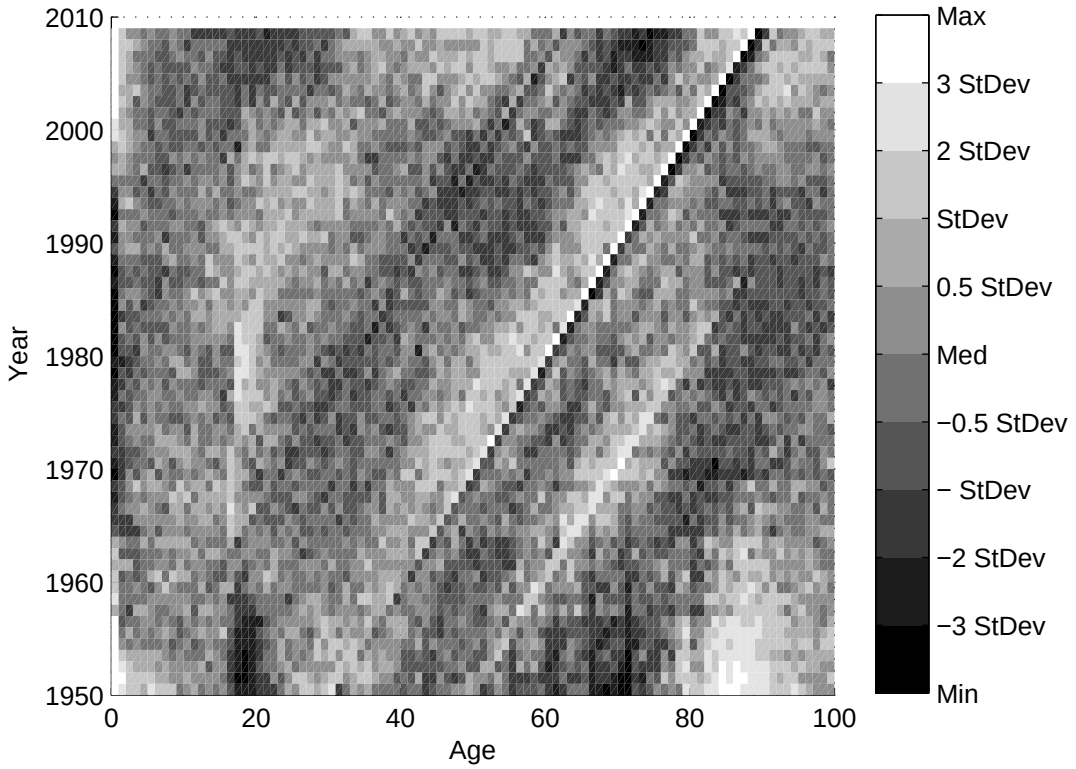


FIGURE 5.5: Heat map of residuals from Stage 3

5.4.5 Stage 4 onwards - Additional age/period terms

The format of the GP from Stage 4 onwards follows the same pattern as for Stages 1, 2 and 3: choose an appropriate functional form for the age term in order to capture the main effect revealed by the non-parametric $\beta_x\kappa_t$ term.

We have already dipped into our toolkit of age functions, most notably by using the two-parameter Gaussian function at Stage 3. Stage 4 and onwards require us to have a far greater range of functions available in the toolkit that we can potentially use. Appendix 5.A contains a list of the parametric functions considered in this analysis.

Figure 5.6 shows plots of the final fitted age functions $f^{(i)}(x, \theta^{(i)})$ and trends $\kappa_t^{(i)}$ for $i = 4, 5, 6, 7$. It is useful to note that the order of discovery of these functional forms

provides a natural order of importance for the age terms.

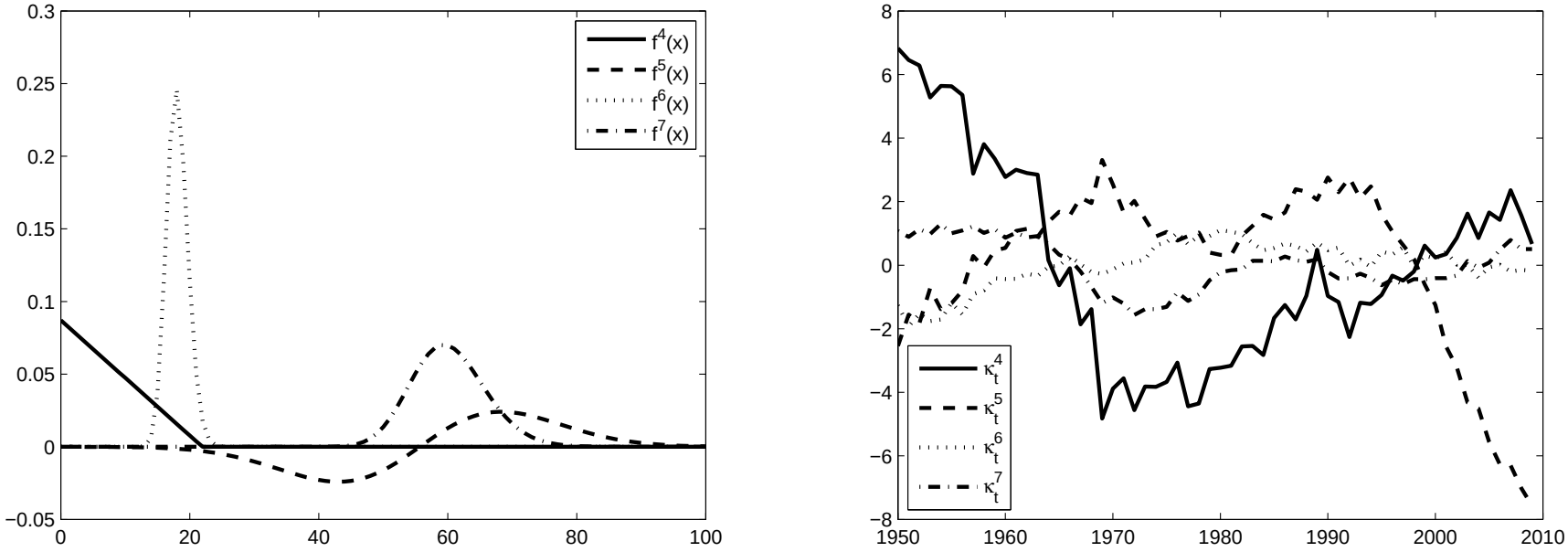
The age functions we have fitted are:

- Stage 4: a broken linear function similar to the payoff of a put option, which we can associate with childhood mortality rates;[9]
- Stage 5: a Rayleigh function, which we associate with the postponement of deaths from late middle age to old age that results from medical improvements over the past 60 years;
- Stage 6: a log-normal function centred on ages 18-19 which we associate with the peak age of the accident hump; and
- Stage 7: a normal function centred on ages 55 to 65 which may be associated with the major causes of death in late middle age, such as lung cancer and coronary heart disease and the efforts made to tackle them.

The residual heat map for Stage 7 (Figure 5.7) is dominated by the diagonal lines representing the cohort effects which have been excluded from the model so far. This might lead us to conclude that we have extracted all of the important age/period effects from the data. This is confirmed by adding a further exploratory non-parametric term to the model. Whilst the resulting BIC for the model does increase, there is little structure to the β_x fitted (shown in Figure 5.8a) except for the periodic pattern at high ages which is clearly trying to capture a series of cohort effects.[10] We therefore conclude that, for UK male data over the sample period, there are seven distinct age/period effects in the data.

[9] This function can be thought of as a very simple linear spline with a single knot, similar to those used as basis functions in Aro and Pennanen (2011). More complex splines could also be considered as part of the toolkit of age functions.

[10] We have tested whether the use of an indicator function at age 18 or a narrow, triangular "spike" function centred on this age would improve the goodness of fit. However, when using the BIC which penalises for excessive parametrisation, the use of these functions did not improve the fit of the model. The use of an indicator function also leads to mortality rates at age 18 being fit perfectly which does not accord with our desire for parsimony and may lead to discontinuous mortality rates which are not biologically reasonable.



(A) Age functions

(B) Period functions

FIGURE 5.6: Age and period functions for Stages 4 to 7 of the general procedure

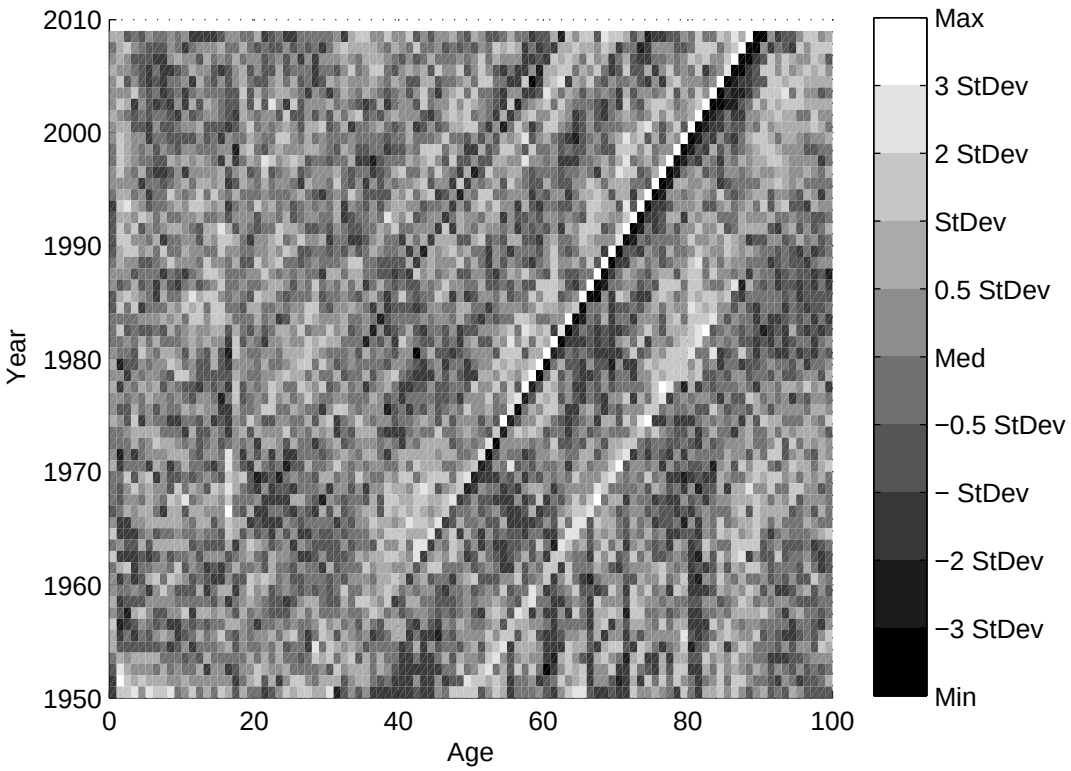


FIGURE 5.7: Heat map of residuals from Stage 7

5.4.6 Stage 8 - Cohort term

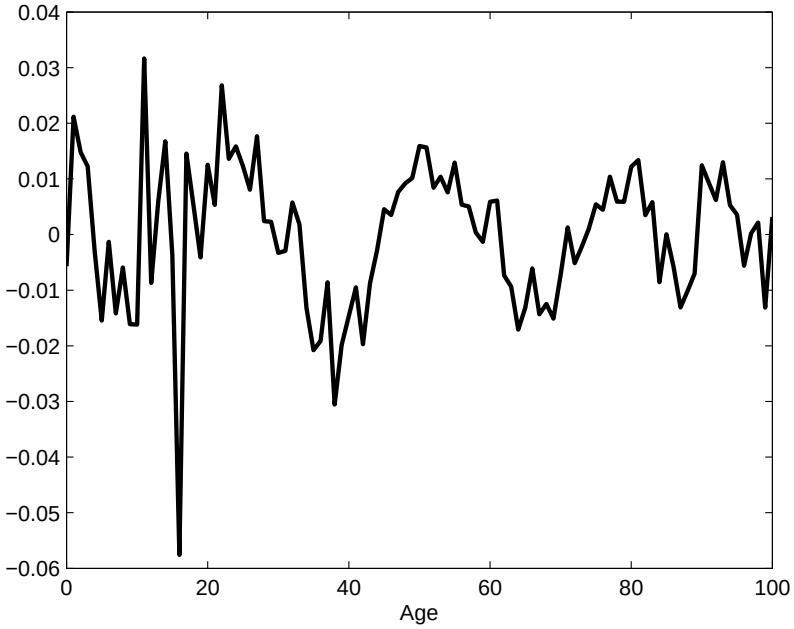
The final stage is to add the cohort parameters γ_{t-x} to yield the final model

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{7} f^{(i)}(x;\theta^{(i)})\kappa_t^{(i)} + \gamma_{t-x}$$

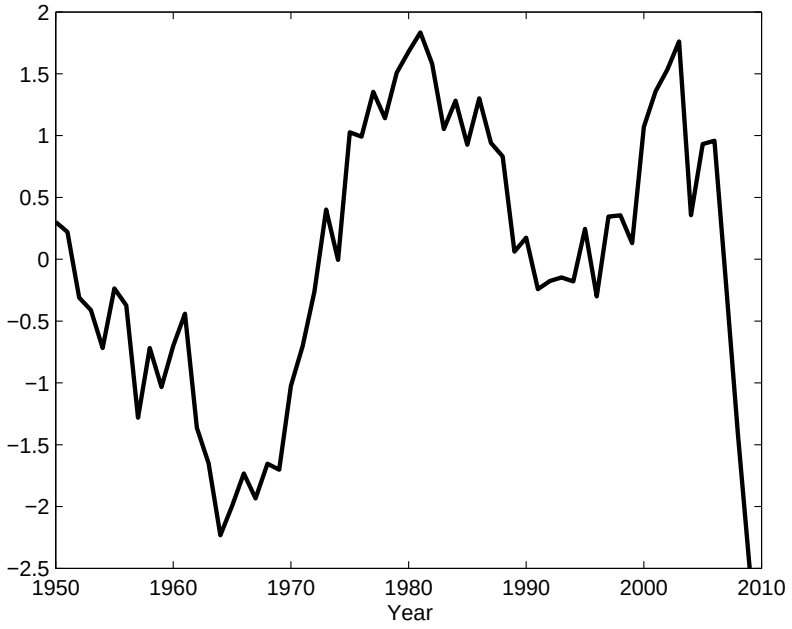
Due to the limited number of observations on very early and late cohorts, we do not estimate cohort parameters in the first and last ten years of birth. Instead, we linearly interpolate these to zero for smoothness. The final model gives the cohort parameters shown in Figure 5.9. Adding a cohort term to the model also creates additional issues with the identifiability of the parameters, which are solved by applying extra identifiability constraints.[11] The full set of identifiability constraints required by the final model produced by the GP is given in Appendix 5.A.

From this, we can identify the major features of interest and can try to relate them to the life histories of the affected cohorts. Most obviously, there is a clear discontinuity between years of birth 1918 and 1919. This may relate to the impact of the influenza

[11] This issue is discussed in Chapter 4.



(A) Age function



(B) Period function

FIGURE 5.8: Non-parametric age and period functions at the end of Stage 7 of the general procedure

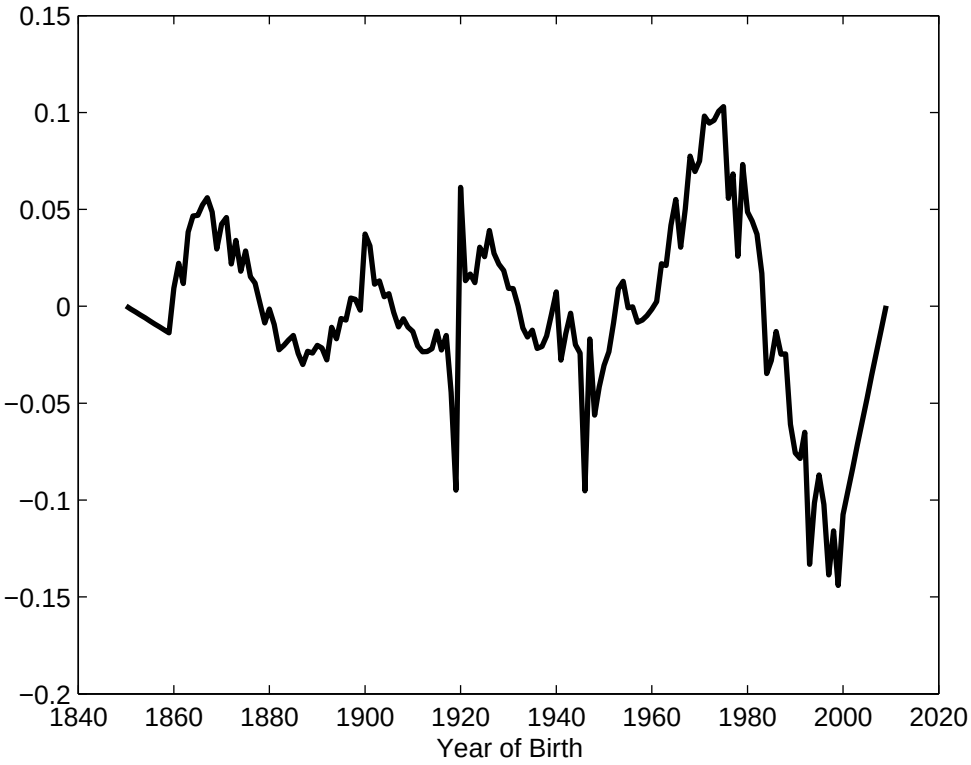


FIGURE 5.9: γ_{t-x} cohort effects from Stage 8 of the general procedure

epidemic that year. Alternatively, it could be a data artefact caused by a flood of births after the First World War distorting the assumptions used to construct exposures to risk (for a discussion, see Richards (2008)). Following this is the decline in cohort mortality observed in Willets (1999, 2004) and discussed in Murphy (2009) relating to the "golden cohort" of individuals born in the late 1920's and early 1930's. We also observe a further (although smaller) discontinuity between 1945 and 1946 relating to the end of the Second World War, strengthening the data artefact argument presented in Richards (2008). We are unsure what demographic significance the excess cohort mortality observed for years of birth between 1960 and 1980 has. These are individuals currently aged between 30 and 50 and therefore we have limited mortality experience data for them and so any attempt at assigning demographic significance is somewhat speculative. However, this feature is robust when adjusting the range of the data for the model and when additional age/period terms are added. This feature will be significant for projecting mortality rates if this excess mortality is continued later into life. Finally, we observe a distinct cohort effect for individuals born around the year 1900 (which again is robust to the model and data specification). This may be due to the formative impact of experience during the First World War as young men and the lifetime health effects this may have induced.

5.5 Testing the final model

Our final model consists of the seven age period terms described in Table 5.1 plus terms for the static life table α_x and the cohort parameters γ_{t-x}.

Term	Description	$f^{(i)}(x) \propto$	Demographic Significance
1	Constant	1	General level of mortality
2	Linear	$x - \bar{x}$	"Gompertz slope", rectangularisation
3	Normal	$\exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right)$	Young adult mortality
4	"Put option"	$(x_c - x)^+$	Childhood mortality
5	Rayleigh	$(x-\hat{x})\exp\left(-\rho^2(x-\hat{x})^2\right)$	Postponement of old age mortality
6	Log-normal	$\frac{1}{x}\exp\left(-\frac{(\ln(x)-\hat{x})^2}{\sigma^2}\right)$	Peak of accident hump
7	Normal	$\exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right)$	Late middle / old age mortality

TABLE 5.1: Age/period terms in the final model

Figure 5.10 shows (on a logarithmic plot) the contribution each of these terms makes to improving the goodness of fit (measured by the BIC) of the model. It can be seen that the majority of the improvement in goodness of fit comes from the first three age/period terms. However, the other terms (as well as being statistically and demographically significant) are still important in describing genuine structure in the data such as the level of inequality in lifespan in the population, described by measures such as the entropy or Gini coefficient of the life table (for instance, see Shkolnikov et al. (2003)). Without them, the cohort term - as the final catch-all term added to the model - would attempt to capture this structure, leading to it being wrongly specified and generating inaccurate and implausible forecasts of mortality rates when projected.

Our final model should, ideally, satisfy the desirable properties relating to the adequacy and goodness of fit of the model discussed in Section 5.2. Specifically

1. it should provide a good and parsimonious fit to the data (which should have been achieved through the model fitting procedure);

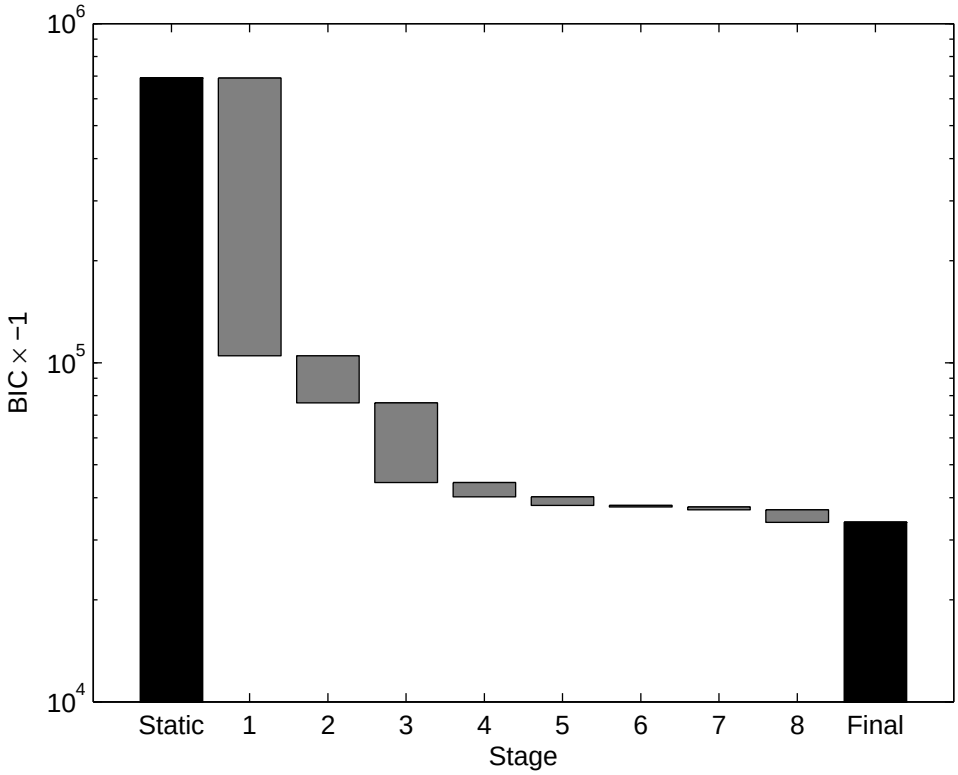


FIGURE 5.10: Improvement in goodness of fit at different stages of the general procedure

2. it should extract all of the significant structure from the data, leaving residuals which are independent and identically distributed; and

3. it should give parameter estimates which are robust to small changes in the data.

To test for structure within the standardised deviance residuals, we extend the procedures in Dowd et al. (2010c). We first plot the heat map shown in Figure 5.11. This shows an apparent lack of any major age/period or cohort features and there are very few "hot" and "cold" regions or clusters in the plot. We then calculate the sample moments of the residuals which are shown in Table 5.2. With large exposures and death counts and assuming the residuals have constant variance, we can use an approximation to assume that they are $N(0,1)$ variables under the null hypothesis and so use the Jarque-Bera statistic to test for this.

The critical statistic for the Jarque-Bera test at 95% is 5.99, whilst at 99% it is 9.21. This means that we decisively reject the assumption of normality for the standardised deviance residuals. Next, we consider the correlations of the residuals with those adjacent

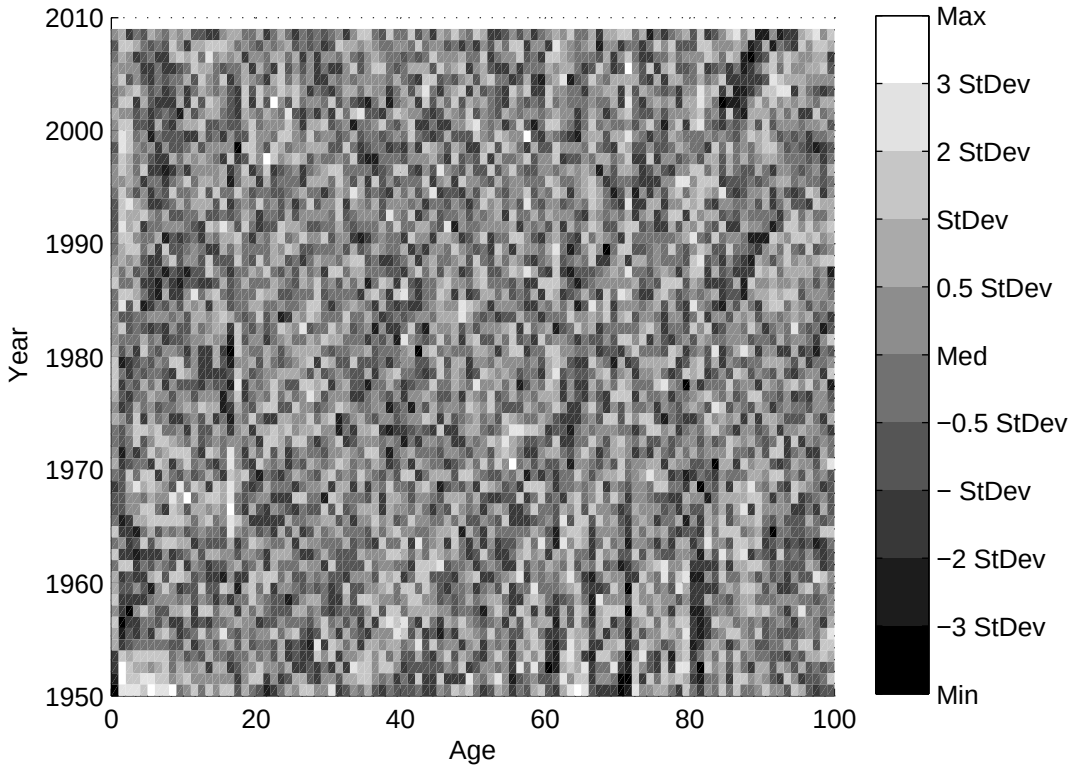


FIGURE 5.11: Heat map of residuals from Stage 8

	Residual mean	Standard deviation	Residual skewness	Residual kurtosis	Jarque-Bera statistic
General procedure	-0.01	0.94	-0.03	3.38	37.70
Lee-Carter	-0.02	0.98	0.47	9.75	11,700
PCA	0.00	0.94	0.06	3.26	21.25

TABLE 5.2: Properties of the residuals from Stage 8 of the general procedure and the Lee-Carter and PCA models

in the age and period directions, i.e.

$$\rho_x^X = corr(\epsilon_{x-1,.}, \epsilon_{x,.})$$
$$\rho_t^T = corr(\epsilon_{.,t-1}, \epsilon_{.,t})$$

Figure 5.12 shows the plot of these correlations against age and year and the relevant statistics if we test against the null hypothesis of independence (a two-tailed test at 95% significance) for the final model from the general procedure. Clearly, the hypothesis of independence is not supported overall. Testing these jointly (i.e., as a series of independent binomial trials where the probability of failure is 5% under the null) confirms the lack of independence in both the age and period directions at the 99% level.

This lack of normality and independence should be investigated further. In practice, this may be due to isolated outliers (often caused by data errors) or due to structural changes within the data. This would cause the variance of the residuals to change with age or time. Plots of the residuals from the model against age, period and cohort (not shown) indicate that there are no extreme outliers that would need to be investigated and that the variance of the residuals is roughly constant. Therefore, it is probable that there is unexplained structure remaining within the data which is not captured by the model. However, comparing these results to those from the PCA model and other models such as the Lee-Carter model show that the GP gives results which are at least as good as those from alternative mortality models.[12]

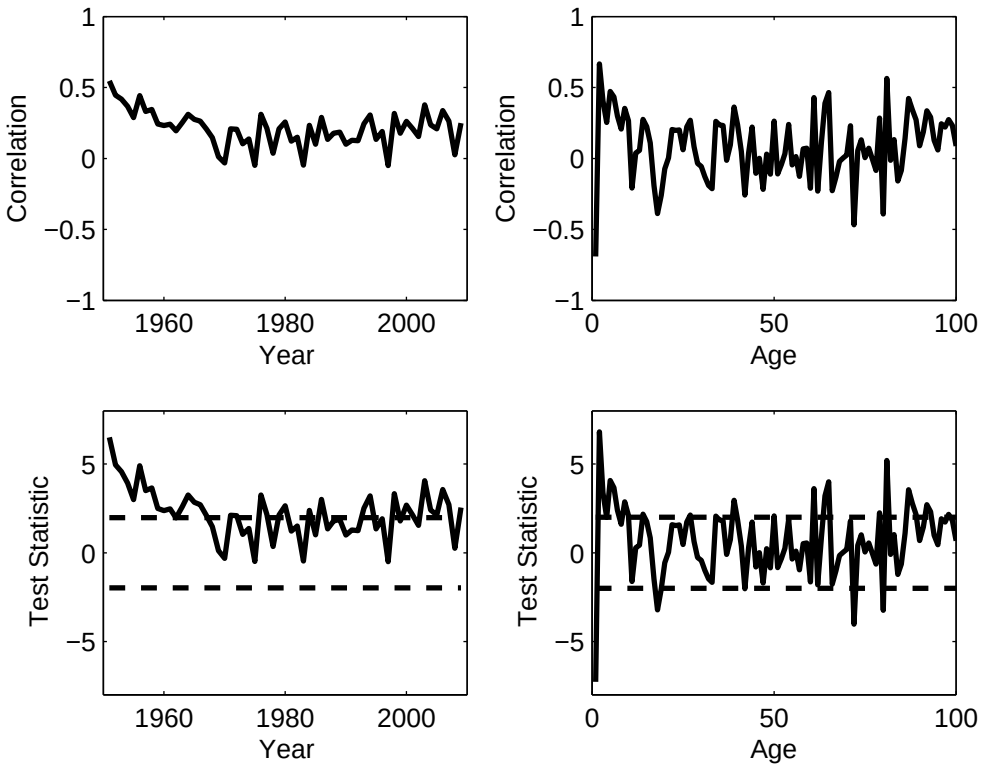


FIGURE 5.12: Correlations and tests statistics for residuals from the general procedure

We also perform a number of tests of the robustness of the model to changes in the data. These include:

1. Fitting the model to different periods of data by increasing the start date sequentially from 1950 to 1980;

[12]We will compare the relative performance of alternative mortality models in Section 5.6.

2. Bootstrapping the standard deviance residuals using a method based on the procedure of Koissi et al. (2006) to test the extent of parameter uncertainty; and

3. Removing ages and years from the data by setting their weights to zero to test that none of the age/period functions are overly sensitive to specific ages and years.

The first of these tests is based on the procedure in Cairns et al. (2009). Graphs of the fitted parameters (not shown but available from the authors) indicate that the model fits similar patterns for the evolution of the different $\kappa_t^{(i)}$ period functions and slowly varying age functions as the age range of the data is changed.

The second robustness test we perform is to look at parameter uncertainty under residual bootstrapping. Standard bootstrapping techniques, such as that implemented by Koissi et al. (2006) were developed for use with the Lee-Carter model and assume that the residuals from the model are independent. However, this assumption is not valid.[13] Nevertheless, for simplicity, we implement an approach based on this method of residual bootstrapping in order to test our final model for parameter uncertainty. This method samples randomly from the fitted residuals and adds them to the fitted mortality surface to generate artificial death counts, to which the model is refitted to generate new parameter estimates. In this fashion, the degree of parameter uncertainty can be ascertained. The plots in Figure 5.13 depict fan charts (see Dowd et al. (2010a)) showing the 90% confidence interval for the period and cohort parameters produced by this bootstrapping procedure using 1,000 simulations. As can be seen, the underlying pattern of the parameters remains unchanged and there is no evidence to suggest that any terms are not significant when allowance is made for parameter uncertainty. The age functions are not shown, but these are considerably more robust to the effect of parameter uncertainty than the period and cohort effects.

As a final test of the model, we systematically remove ages and years from the data by setting their weights to zeros and then refitting the parameters. This tests if any of the fitted functions are overly sensitive to the specific rows or columns of the data grid, and the model's ability to interpolate sensibly for missing data. Figures 5.14 and 5.15 shows the impact of this analysis on the cohort parameters γ_{t-x} and on the age/period terms $f^{(6)}(x)$ and $\kappa_t^{(6)}$.[14] As can be observed, while removing specific ages and years can distort the cohort parameters at the end of the range of data, it does not substantially

[13] More recently, stratified (see D'Amato et al. (2011)) and block-bootstrapping (see Liu and Braun (2010)) procedures have been used, as have those based on geo-statistical techniques which look at the correlation structure across residuals (see Debón et al. (2008, 2010)).

[14] This age/period term was chosen as the most specific age function fitted and therefore probably the most susceptible to uncertainty under this analysis.

affect those estimated across more data points in the centre of the range. $\kappa_t^{(6)}$ is also robust under this analysis.[15]. We are therefore satisfied that our final model is robust under small changes to the data.

5.6 Comparison with alternative models

The model produced by the GP in Section 5.5 had some unexplained structure according to our analysis of the residuals. How serious a problem is this? Perhaps the best way to answer this question is to compare the model from the GP with some alternative mortality models: the LC model (as the most widely used mortality model) and a method based on principal component analysis which extends the Lee-Carter approach with multiple age/period and cohort terms.

The LC model, introduced in Lee and Carter (1992) has subsequently been much studied, developed and extended, most notably in the work of Lee (2000), Brouhns et al. (2002a), Booth et al. (2002), Renshaw and Haberman (2003b, 2006) and Hyndman and Ullah (2007). It has rapidly become the benchmark mortality model against which others are compared (for instance in Cairns et al. (2009) or Plat (2009a)) and so is a natural starting point for comparing the model produced by the GP against. However, it is a relatively simple model with only one age/period term and no cohort term, and so we would expect the GP to give significantly better fits to the data.

The singular value decomposition used to fit the model to data in Lee and Carter (1992) is a particular implementation of principal component analysis (PCA) - see Huang et al. (2009) for more details. It is therefore the natural extension of the Lee-Carter methodology capable of giving multiple age/period terms. It finds age and period functions that explain the maximum amount of variance (across the period dimension) in the model. PCA has long been used in the study of mortality rates: for example Wilmoth (1990) used it to detect higher order age/period functions, Booth et al. (2002) and Renshaw and Haberman (2003b) both proposed its use to extend the Lee-Carter model with additional age/period terms and the models of Hyndman and Ullah (2007) and Yang et al. (2010) used it directly to fit multiple age/period effects. However, it cannot directly find cohort effects. Therefore a direct comparison of PCA with our model is not appropriate.

[15] Corresponding graphs for the age functions and other period functions, not shown here, also show considerable robustness.

In order to compare procedures, we use a method similar to that used in Wilmoth (1990). We first use PCA to find age/period functions for $\ln(\mu_{x,t})$ in the absence of cohort effects. We then add a cohort effect to the underlying model and use the PCA age/period effects as the starting point when maximising the Poisson log-likelihood using the algorithms in Appendix 5.A. This process is repeated for different numbers of age/period terms and the model with the highest BIC selected for comparison against our final model.

5.6.1 Results

Table 5.3 compares the three models and shows the goodness of fit to our dataset. The LC is a single factor model and so it is unsurprising that the other two models give considerably better fits to the data, although at the cost of a far greater number of parameters. The PCA method also requires substantially fewer age/period terms to achieve a very similar goodness of fit to the model produced by the GP. Because each of these age functions has approximately one hundred free parameters compared with a maximum of two using the GP, this does not result in a more parsimonious model, however. Further, as we are primarily interested in the evolution of mortality rates over the period, we consider that it is desirable to have a high proportion of the parameters relating to the period and cohort effects of interest. This is not the case in the PCA model.

Model	No. A/P terms	No. free parameters	Log-likelihood	BIC
General procedure	7	679	-3.09×10^4	-3.38×10^4
Lee-Carter	1	259	-5.13×10^4	-5.25×10^4
PCA	3	735	-3.07×10^4	-3.39×10^4

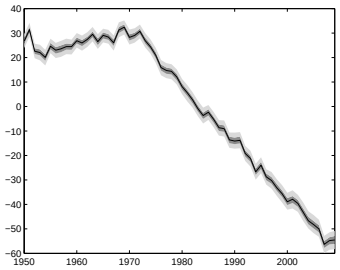
TABLE 5.3: Goodness of fit for the different models

Figures 5.16 and 5.17 show the age and period functions for the GP and PCA procedure - the age and period functions for the LC model are the same as the non-parametric terms shown in Figure 5.2. We find it difficult to assign demographic significance to the age functions in the LC and PCA models. The cohort parameters for the GP and PCA models are shown in Figure 5.18 - there is no corresponding plot for the LC model due to the absence of a cohort term. Here it is worth noting the similarities as well as the differences in the fitted parameters. Both approaches detect the discontinuities after the First and Second World Wars and the increase in cohort mortality for years of birth around 1900 and between 1960 and 1980.

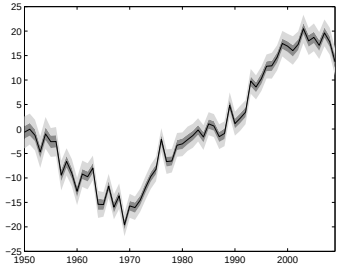
However, there are substantial differences in both the magnitude and the pattern of cohort parameters. Cohort effects for the GP are less pronounced than those from the PCA procedure. In addition, the PCA model fails to find a sustained decrease in cohort mortality for the "golden cohort" discussed previously. Most seriously, there appear to be large cohort effects at the beginning and end of the range of years of birth which are not explainable demographically. We believe that these effects are trying to compensate for the second and third age functions in the PCA model, which do not tend to zero at high ages (as shown in Figure 5.17a). This has very serious effects when these models are projected into the future. We therefore believe that the cohort parameters produced by the GP are more biologically reasonable and demographically significant than those fitted by the PCA procedure.

Table 5.2 above shows the moments and results of the Jarque-Bera tests on the residuals for the three approaches. We note that none of the three models tested give normally distributed standardised residuals, although the residuals from the GP and PCA models come considerably closer than those from the LC model.

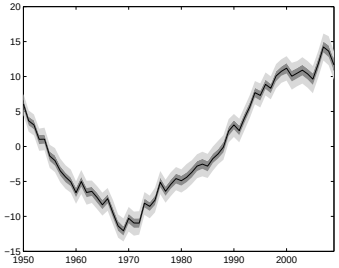
We also compare plots of the residual heat maps in Figure 5.19 and test for correlation amongst the standardised deviance residuals in Figure 5.20 from the Lee-Carter and PCA models in Figure 5.20 - comparable plots for the GP are shown in Figures 5.11 and 5.12 respectively. The heat maps for the Lee-Carter and PCA models shows obvious clusters in the fitted residuals, indicating that there is still substantial structure remaining in the residuals of the PCA model. The LC residuals in particular show the clear need for a cohort term to capture the impact of the cohorts born after the First and Second World Wars. The PCA model yields residuals which are closer to normality than the GP, although they still do not pass the Jarque-Bera test. The correlations across residuals from the PCA procedure are higher than from the GP. Probably this is due to the smaller number of age/period terms. However, adding additional terms to the PCA model results in worse BICs and therefore will not improve the goodness of fit. This reinforces the conclusion that there is still structure in the data which is not adequately captured by the PCA model.



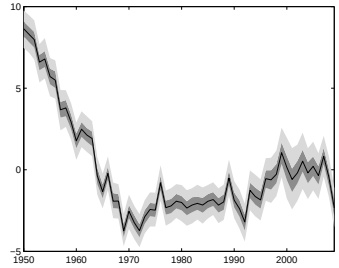
(A) $\kappa_t^{(1)}$



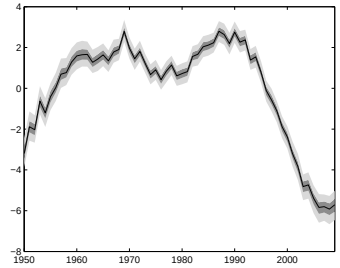
(B) $\kappa_t^{(2)}$



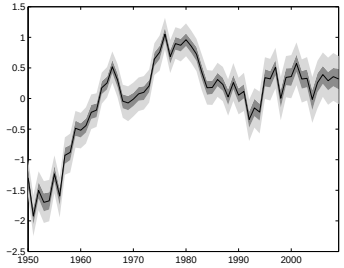
(C) $\kappa_t^{(3)}$



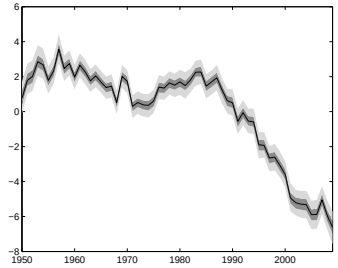
(D) $\kappa_t^{(4)}$



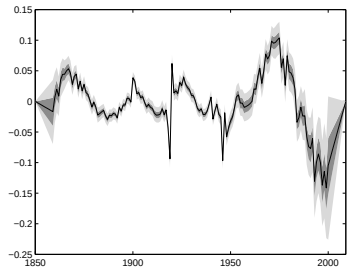
(E) $\kappa_t^{(5)}$



(F) $\kappa_t^{(6)}$

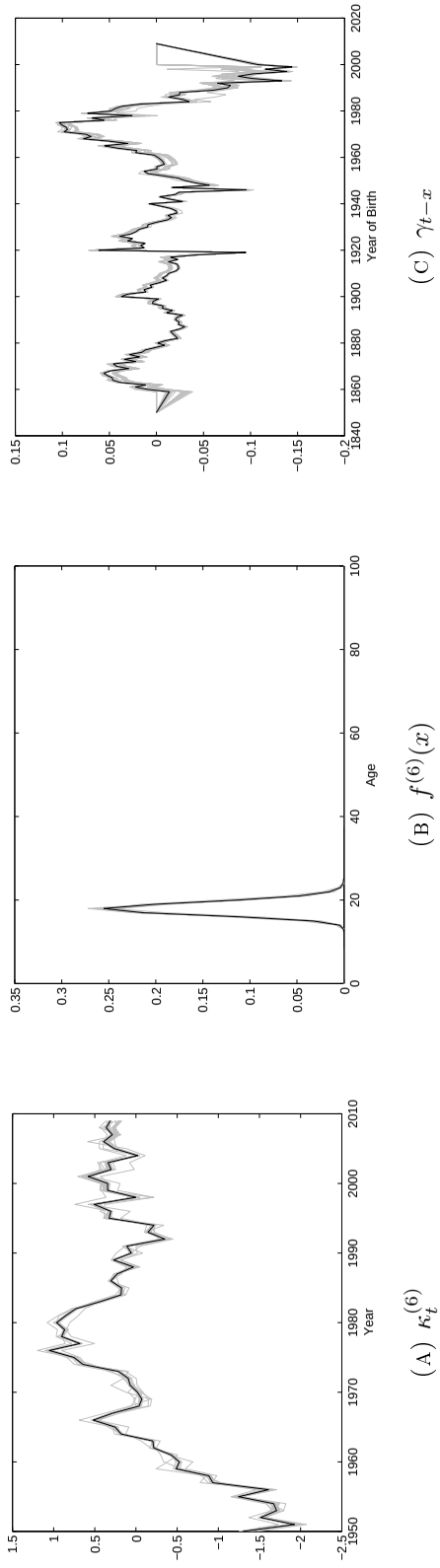


(G) $\kappa_t^{(7)}$



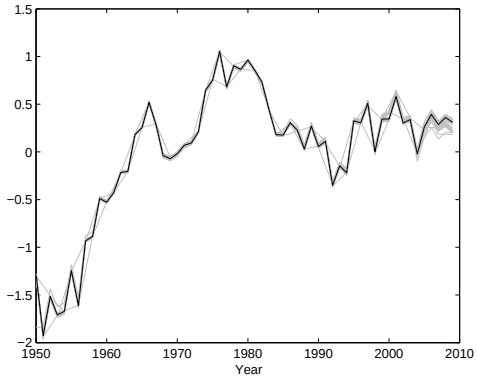
(H) γ_{t-x}

FIGURE 5.13: Parameter uncertainty due to residual bootstrapping

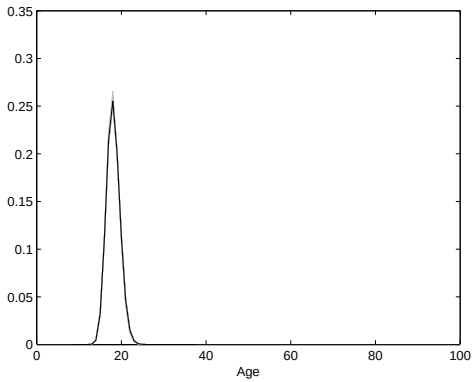


(A) $\kappa_t^{(6)}$ (B) $f^{(6)}(x)$ (C) γ_{t-x}

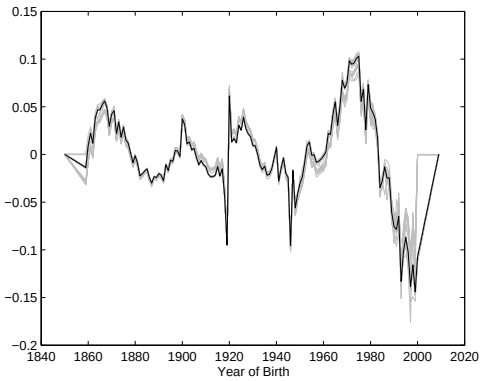
FIGURE 5.14: Parameter uncertainty due to removal of one age of data



(A) $\kappa_t^{(6)}$

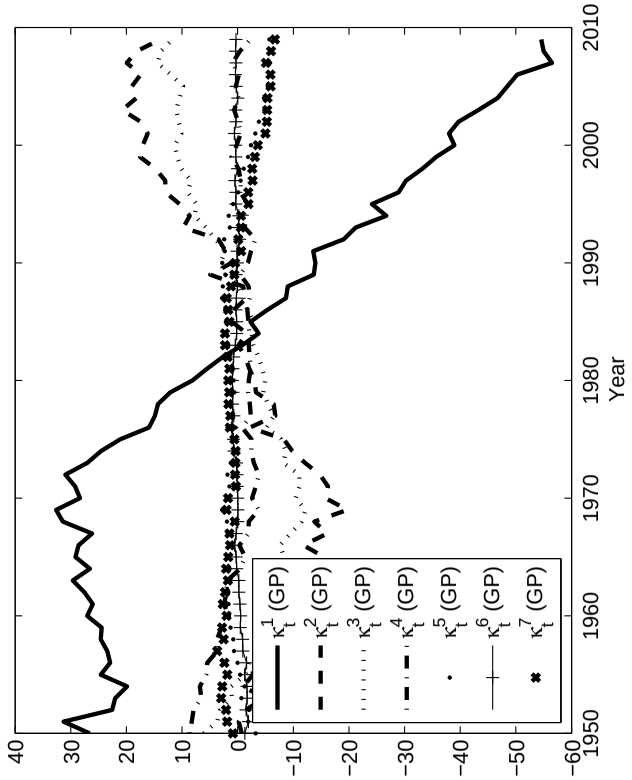


(B) $f^{(6)}(x)$

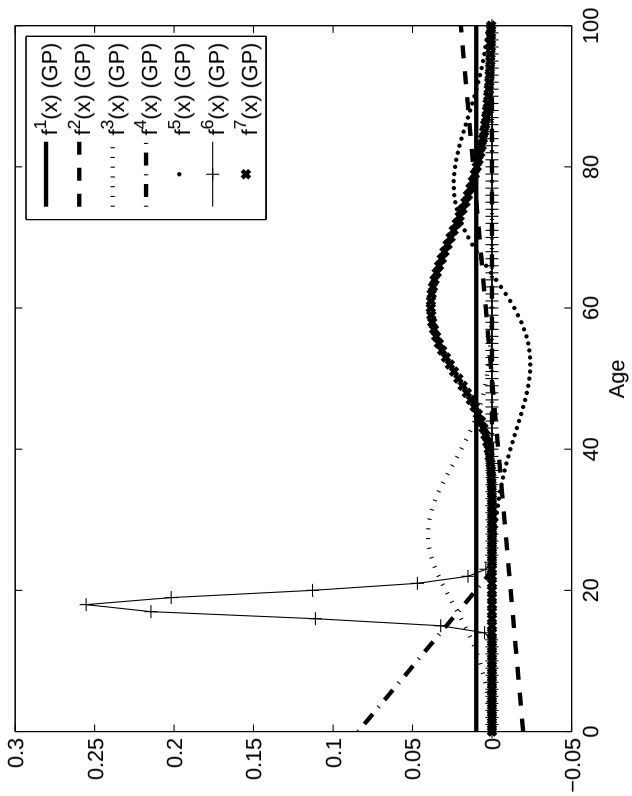


(C) γ_{t-x}

FIGURE 5.15: Parameter uncertainty due to removal of one year of data

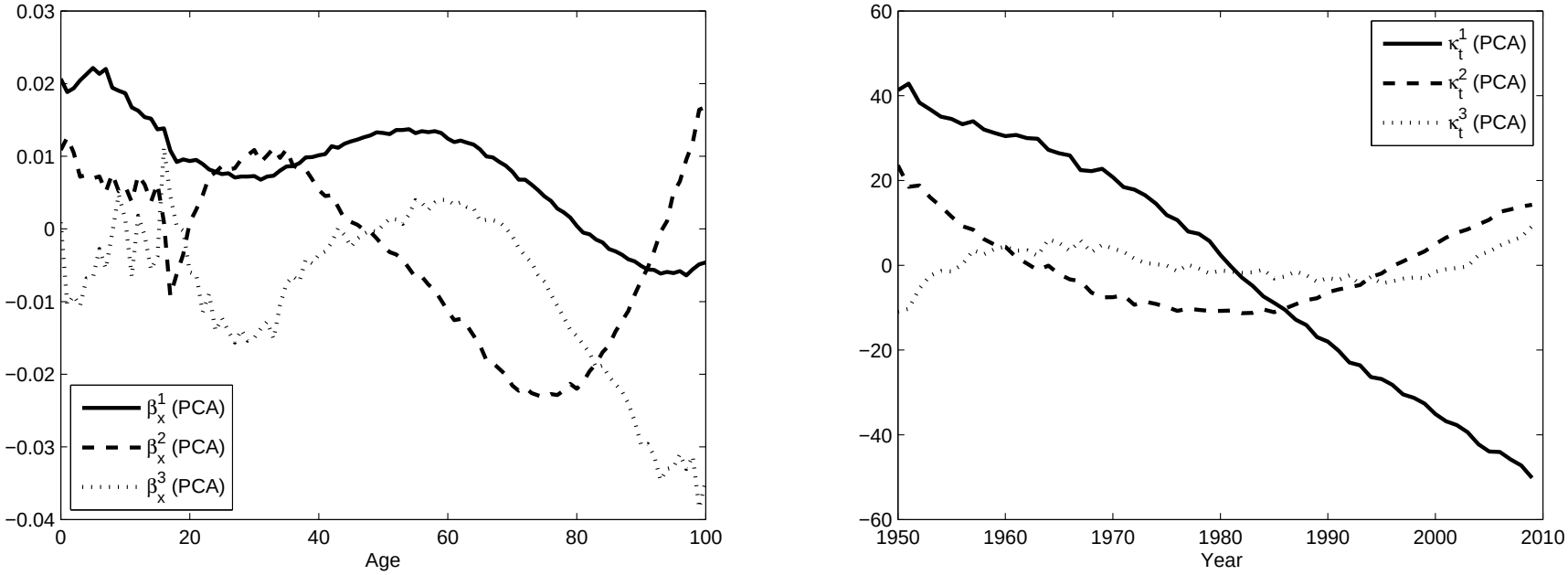


(B) Period functions



(A) Age functions

FIGURE 5.16: Age and period functions for the general procedure



(A) Age functions

(B) Period functions

FIGURE 5.17: Age and period functions for the PCA model

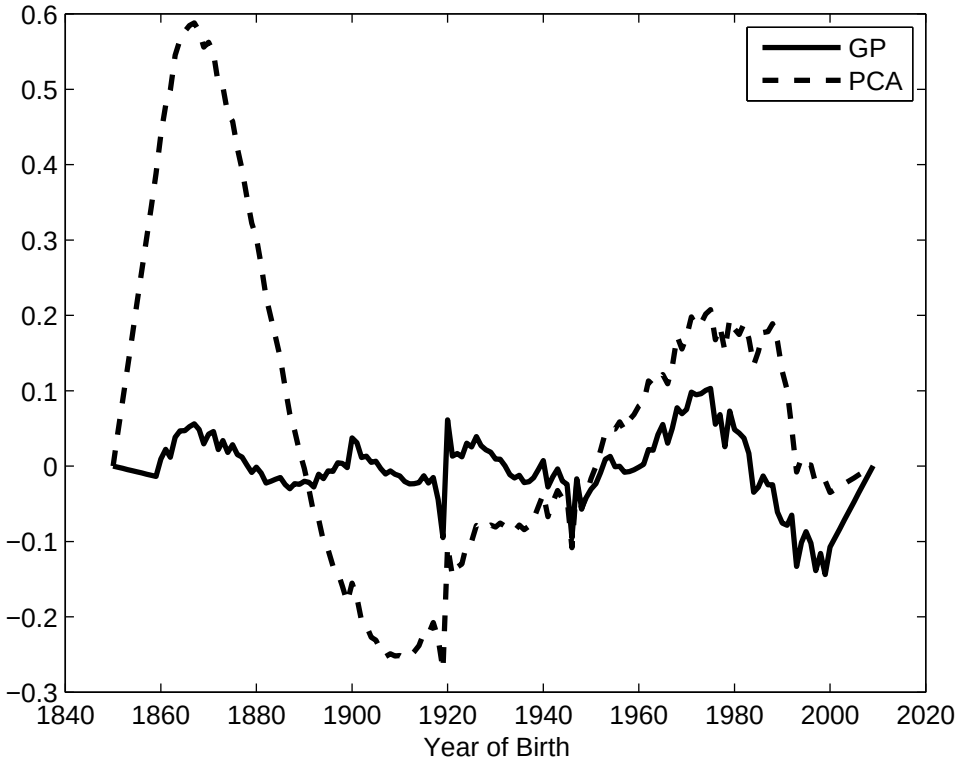
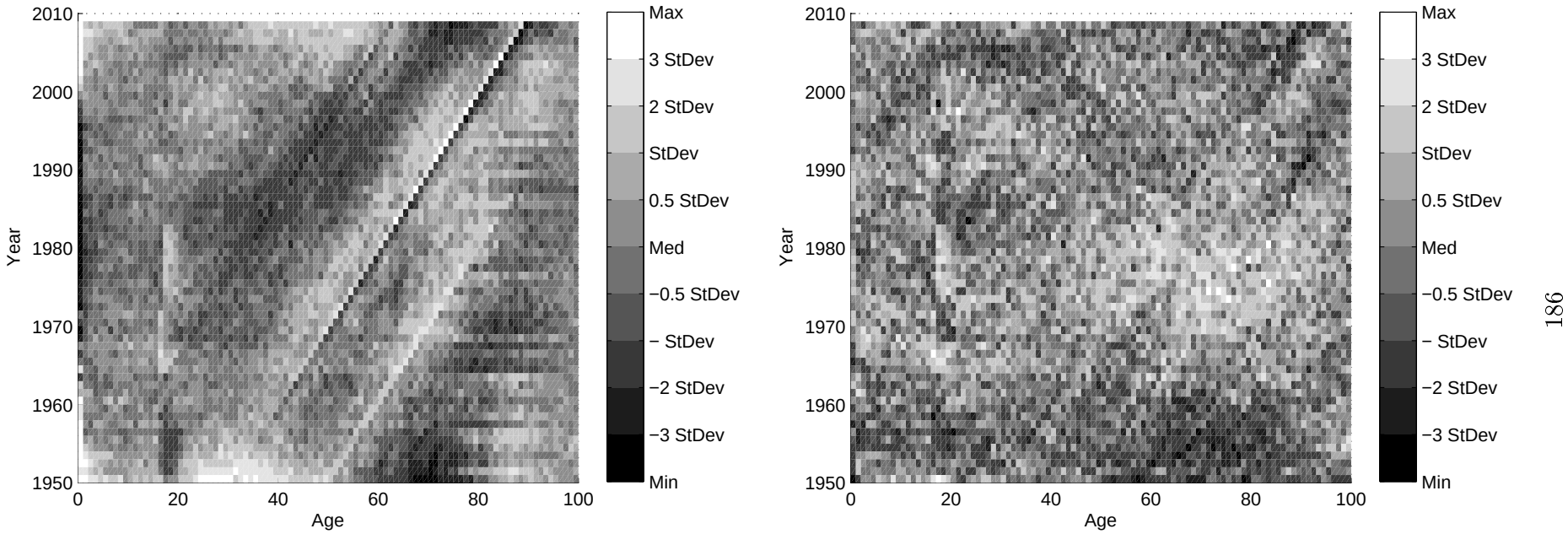


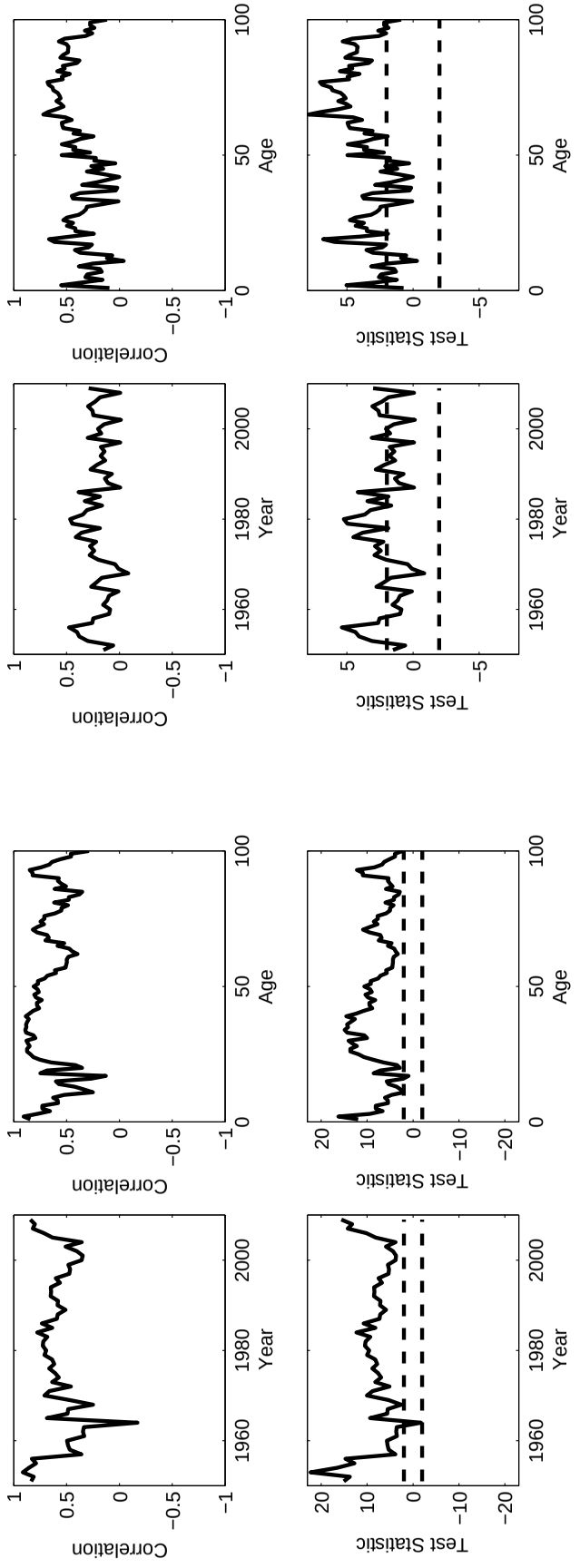
FIGURE 5.18: Cohort parameters for the GP and PCA models



(A) Lee-Carter

(B) PCA

FIGURE 5.19: Residual heat maps for the Lee-Carter and PCA models



(A) Lee-Carter (B) PCA

FIGURE 5.20: Residual correlations across age and period for the Lee-Carter and PCA models

5.7 Conclusions

As the level of interest in longevity risk increases, it becomes increasingly important to be able to construct more sophisticated mortality models reliably and robustly. These will need to capture most of the identifiable structure in mortality rates within the data - which calls for more terms - but to do so with the smallest number of free parameters - which calls for parsimony. Where cohort effects are believed to be real and important, they will need to be captured by the model. However, they must also be clearly distinguished from age/period effects in order that they can be projected correctly. This, in practice, means that all the significant age/period effects must be identified before any attempt is made to estimate the cohort effect. Finally, terms within the model should be capable of being associated with underlying biological or social processes. This requires judgement to be used to guide their projection and aid their communication with other, non-technical, stakeholders who are subject to longevity risk and wish to understand the implications.

In this chapter, we have introduced a new, general procedure for constructing mortality models. The general procedure is driven by forensically examining the data to provide evidence for the selection of each and every term in the final model produced. We believe this improves the goodness of fit of the model parsimoniously and with demographic significance. We have applied the general procedure to a specific dataset, associated each term generated with an underlying demographic and/or socio-economic factor for the population being modelled, analysed the residuals to confirm that there is no identifiable structure remaining in the data which is not captured by the model, and compared the results with those from other methods of constructing mortality models.

The general procedure requires the modeller to engage intelligently with the data and make various subjective decisions in its implementation. It is not a "black box" algorithm which can be deployed mechanically on various datasets, but rather requires a substantial investment of time to understand the underlying forces driving mortality within the population of interest and how these forces can be represented mathematically. But far from this being a disadvantage, we would argue that our approach accords perfectly with good model building practice, which seeks to move beyond a purely algorithmic approach in order to understand better the underlying structure of the data.

In conclusion, we believe that the general procedure is capable of producing models which are in accordance with the desirability criteria of adequacy of fit to the data, demographic significance, parsimony, robustness and completeness (by including sufficient terms to cover all ages and cohorts).

However, we are aware that in order to be practically useful, a good fit to historical data needs to be accompanied by the ability to use the model to make reliable forecasts of future mortality rates. Projecting models with multiple age/period and cohort terms consistently is a difficult problem as the historical time series are often highly correlated and display curvature, outliers or subtle trend changes which need to be accommodated (as have been described in Li and Chan (2005), Li et al. (2011) and Coelho and Nunes (2011)). We address these issues in Chapters 6 and 8.

5.A Appendix: Algorithms and toolkit of function

In order to implement the general procedure, we need the ability to introduce new terms to existing models and to fit these to data. At each stage, all parameters within the model are freely estimated (although the values found at previous stages are used as convenient starting points for later stages of the maximisation algorithm). The exception to this is when new non-parametric terms are added to the model and the previously fitted age functions are not re-estimated as this often leads to model instability. As these terms are added purely for exploratory purposes and all parameters will be re-estimated once they are replaced with suitable parametric forms, we do not believe this will have a significant impact on the final model.

As we have central exposures to risk from the Human Mortality Database (Human Mortality Database (2014)), we adopt a Poisson likelihood maximisation approach which enables us to do this quickly and efficiently. This procedure is based on that implemented in Brouhns et al. (2002a) and is described in Algorithm 1 at high level below.

The fitting algorithm used by the general procedure differs from the Brouhns et al. (2002a) method in that the log-likelihood is maximised with respect to each set of parameters sequentially rather than simultaneously. It could be argued that this may lead the algorithm to find local rather than global maxima for the parameter values. In practice, we have not found this to be an issue and believe it can be largely resolved through finding the full set of identification issues for the parameters within the model

Algorithm 1 Algorithm for Poisson likelihood maximisation

1: Set initial starting values and calculate initial log-likelihood
2: **while** Increase in log-likelihood less than threshold value (e.g. 10^{-2}) **do**
3: Maximise log-likelihood with respect to α_x holding all other parameters constant
4: **for** Each age/period term i **do**
5: Maximise log-likelihood with respect to $\kappa_t^{(i)}$ holding all other parameters constant
6: Maximise log-likelihood with respect to free-parameters $\theta^{(i)}$ in age function $f^{(i)}(x;\theta^{(i)})$ or with respect to β_x holding all other parameters constant
7: **end for**
8: Maximise log-likelihood with respect to γ_{t-x} holding all other parameters constant if model contains a cohort term
9: Impose identifiability constraints through use of invariant transformations
10: Calculate updated log-likelihood
11: **end while**
12: Calculate residuals and BIC

(as discussed in Chapters 3 and 4). The maximisation of each set of parameters (i.e. $\xi = \alpha_x,\ \beta_x,\ \kappa_t^{(i)},\ \gamma_c,\ \theta^{(i)}$) is done as per Algorithm 2 below.

Algorithm 2 Algorithm for maximisation of individual parameters

1: Start with values for maximisation passed from parent algorithm
2: **while** Increase in log-likelihood less than threshold value (e.g. 10^{-4}) **do**
3: Calculate first derivative of log-likelihood with respect to parameters $\frac{\partial L}{\partial \xi}$
4: Calculate second derivative of log-likelihood with respect to parameters $\frac{\partial^2 L}{\partial \xi^2}$
5: Update estimate of parameters $\hat{\xi} = \xi - \phi \frac{\frac{\partial L}{\partial \xi}}{\frac{\partial^2 L}{\partial \xi^2}}$
6: Impose identifiability constraints, e.g. on the level of $\kappa_t^{(i)}$, using invariant transformations
7: Update fitted surface $\mu_{x,t}$ and log-likelihood
8: **end while**
9: Return updated parameter estimates, fitted mortality rates and log-likelihood to parent algorithm

This is nothing more than the repeated application of the Newton-Raphson procedure. The parameter $\phi \in (0, 1]$ is a simple scaling which can be lowered to improve the stability of parameter estimates (albeit at the cost of increasing the run time of the algorithm). In most cases, the parameter sets are treated as vectors meaning that $\frac{\partial^2 L}{\partial \xi^2}$ is the Hessian matrix. However, this matrix usually has a diagonal structure (e.g. $\frac{\partial^2 L}{\partial \alpha_x \partial \alpha_y} = 0$ for $x \neq y$) which simplifies the implementation significantly.

Models produced by the GP will not be fully identified and so will require additional identifiability constraints to be robustly estimated. A discussion of the origin and nature of this lack of identifiability and the selection of appropriate identifiability constraints

was given in Chapters 3 and 4. In summary, we impose the following identifiability constraints upon the final model from Stage 8.

$$\sum_t \kappa_t^{(i)} = 0 \quad \forall i \tag{5.7}$$

$$\sum_x |f^{(i)}(x;\theta^{(i)})| = 1 \quad \forall i \tag{5.8}$$

$$\sum_y n_y \gamma_y = 0 \tag{5.9}$$

$$\sum_y n_y \gamma_y (y - \bar{y}) = 0 \tag{5.10}$$

$$\sum_y n_y \gamma_y ((y - \bar{y})^2 - \sigma_y) = 0 \tag{5.11}$$

Not all of these constraints will be applicable at all stages (e.g., the constraints in Equations 5.9, 5.10 and 5.11 will not apply to models without a cohort term) whilst for models with a non-parametric age function, we require the additional constraints below.

$$\sum_x |\beta_x| = 1 \tag{5.12}$$

$$\sum_x \beta_x f^{(i)}(x;\theta^{(i)}) = 0 \quad \forall i \tag{5.13}$$

$$\tag{5.14}$$

We refer to Equations 5.8 and 5.12as the normalisation of the age function. In contrast to some authors (e.g. Haberman and Renshaw (2009)) we do not require that age functions are non-negative. In order to normalise age functions with free parameters $\theta^{(i)}$, we must modify the form of the age function so that $\sum_x |f^{(i)}(x;\theta^{(i)})|$ is not a function of $\theta^{(i)}$. This means that the normalisation scheme in Equation 5.8 holds as $\theta^{(i)}$ is varied when fitting the model. This is usually achieved by multiplying it by a "self-normalisation" function $N(\theta^{(i)})$. This was discussed in greater depth in Chapter 3. Equation 5.13 is only applied in exploratory models with a non-parametric term in order to maximise the distinctness of the age/period terms.

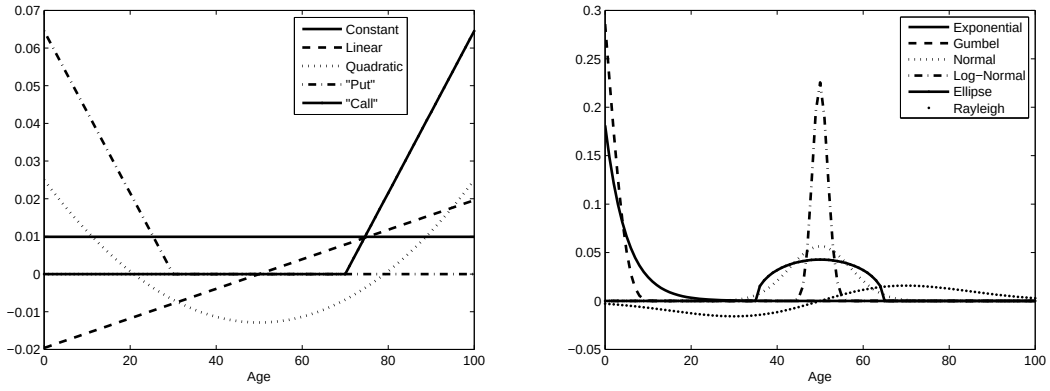
The functions in the toolkit we have developed so far are given in Table 5.4 along with the free parameters they require and the self-normalisation functions $N(\theta^{(i)})$. In this, the age range is assumed to run from age 1 to age X with $\bar{x} = \frac{1}{X}\sum_{x=1}^{X} x$ and $\sigma_x = \frac{1}{X}\sum_{x=1}^{X}(x - \bar{x})^2$. Some of these normalisations are only approximate or are true up to a constant, so it is still necessary to rescale the age functions after applying Algorithm 2 to optimise the value of the free parameters. Similar definitions for $\bar{y}$ and σ_y are

used in Equations 5.9, 5.10 and 5.11.

Name	Function $f(x) \propto$	Normalisation $N(\theta)$	Free Parameters
Constant	1	$\frac{1}{X}$	none
Linear	$x - \bar{x}$	$\frac{1}{\bar{x}(\bar{x}+1)}$	none
Quadratic	$(x-\bar{x})^2 - \sigma_x$	$\frac{1}{12}X(X+2)^2$	none
"Put option"	$(x_c - x)^+$	$\frac{1}{x_c(x_c-1)}$	x_c - pivot
"Call option"	$(x - x_c)^+$	$\frac{1}{(X-x_c)(X-x_c-1)}$	x_c - pivot
Exponential	$\exp(-\lambda x)$	$1 - \exp(-\lambda)$	λ - width
Gumbel	$\exp(\exp(-\lambda x))$	λ	λ - width
Spike	$(x - (x_c - a))I(x_c - a \le x < x_c) +$ $((x_c + a) - x)I(x_c \le x < x_c + a)$	$\frac{1}{a}$	x_c - peak a - width
Normal	$\exp\left(-\frac{(x-\hat{x})^2}{\sigma^2}\right)$	$\frac{1}{\sigma}$	$\hat{x}$ - location σ - width
Log-Normal	$\frac{1}{x}\exp\left(-\frac{(\ln(x)-\hat{x})^2}{\sigma^2}\right)$	$\frac{1}{\sigma}$	$\hat{x}$ - location σ - width
Rayleigh	$(x-\hat{x})\exp(-\rho^2(x-\hat{x})^2)$	$0.5\rho^2$	$\hat{x}$ - location ρ - width^{-1}
Ellipse	$\sqrt{1 - \frac{(x-\hat{x})^2}{a^2}}$	$\frac{2}{a\pi}$	$\hat{x}$ - location a - width

TABLE 5.4: Age functions in toolkit

FIGURE 5.21: Age functions in toolkit



Part II

Projection of Mortality Rates for Single or Multiple Populations

Chapter 6

Consistent Mortality Projections Allowing for Trend Changes and Cohort Effects

6.1 Introduction

The last two decades have seen dramatic changes in the modelling and management of longevity risk, both in theory and in practice. One important change has been the switch from using deterministic models based on expert judgement to stochastic models which extrapolate the observed trends within the data to give probabilistic forecasts of mortality rates.

The extrapolative approach to projecting mortality has the core assumption that there is consistency between the evolution of mortality rates in the past and the future. After all, today is both yesterday's future and tomorrow's past. While it is easy to criticise this assumption as simplistic - as, for example, Gutterman and Vanderhoof (1998) do - and point out the many potential new advances in medicine which may occur in future, it is important to remember that the past also experienced profound innovations that we take for granted today. Revolutions in the provision of healthcare, new epidemics and pandemics, and changes in lifestyle have all affected mortality rates in developed countries since the Second World War. It therefore seems reasonable to use the experience gained from analysing past developments to help us with forecasting the future.

When using extrapolative mortality models, there is a fundamental symmetry between the processes of fitting the model to historical observations in order to estimate parameters, on the one hand, and projecting parameter values to generate future observations, on the other. It is important that the models we use to project mortality genuinely achieve consistency between the past and the future. This ensures that our projections are as similar to those observed in the historical data as possible, both in their central estimates of future mortality rates and in the levels of uncertainty around these estimates. For example, when we fit models to the past, we often see changes in trends in the parameters. For consistency, similar trend changes should also be present in our projections of these parameters in the future. We must also take care when looking at the lifelong features of mortality affecting specific cohorts, since our data only shines a partial light on the life histories of those cohorts with members who are still alive.

We must be aware of the arbitrary choices we make when fitting a model to data, for instance, our choice of which constraints to apply in order to identify the parameters in a model fully. Different choices imply different interpretations of the parameters, but not the fitted mortality rates themselves, and therefore it is important to ensure that these choices do not change our projected mortality rates either. This subject is considered in depth in Chapters 3 and 4 for general age/period/cohort mortality models. In this study, we apply the principles established in those studies to the specific context of the mortality model constructed in Chapter 5 to see how they are applied in practice and the impact they make on the projection of mortality rates.

This chapter discusses the extrapolative approach to projecting mortality and some of the criticisms of it in Section 6.2. It then reviews the mortality model developed in Chapter 5 for men in the UK using the "general procedure" in Section 6.3 and proposes a number of new techniques to project mortality across periods and along cohorts in Sections 6.4 and Section 6.5. These techniques attempt to ensure that there is consistency between the past and the future which is independent of our arbitrary choices made when fitting the model. They are presented in the context of the model developed in Chapter 5, however, they can be applied more generally to any age/period/cohort mortality model, such as those discussed in Chapter 2. Doing so allows us to obtain more accurate forecasts of mortality rates in the short term, but also gives greater variability in our long-term forecasts. In Section 6.6, we show this by using a backtesting exercise to demonstrate the improvements in short-term predictive power and then demonstrate how standard projection methods may understate both the expected values and the riskiness of annuities in payment for example. An additional benefit of these new techniques is more effective risk management, as traditional techniques may understate the risks in

providing long-term benefits.

6.2 The extrapolative approach to projecting mortality rates

The extrapolative approach to projecting mortality analyses the patterns in the evolution of mortality rates statistically and then uses time series methods to project these into the future. It therefore has, as a central assumption, that there is consistency between the past and the future. As Booth (2006, p. 550) said

> Extrapolative methods are essentially atheoretical; the only assumption is that the future will be (in some sense) a continuation of the past. This is their strength, but it is also their fundamental weakness: historical patterns may not be the best guide to the future, notably because changes in the trend, or structural changes, may be missed. Extrapolative methods make no use of exogenous variables: they do not incorporate current knowledge about actual and prospective developments in relevant areas such as medicine and new diseases, lifestyles and the economy.

This embodies the central criticism of the extrapolative method; that a failure to understand and incorporate information regarding medical progress and socio-economic factors makes extrapolative projections unsuitable, as discussed in Gutterman and Vanderhoof (1998). A lot of research has been conducted into analysing these exogenous causes and their impact on mortality rates, for instance in Manton et al. (1980), Ruhm (2000, 2004), Reichmuth and Sarferaz (2008), Gaille and Sherris (2011) and Hanewald (2011). However, it is fair to say that we are still a long way from truly understanding these underlying factors. As stated by Andreev and Vaupel (2006): *"Cause-specific forecasts are of less benefit to long-term forecasts than they are to the short-term variant, however, due to the current lack of knowledge about disease etiology and about the factors underlying mortality trends in the distant future."* A similar point is made in Continuous Mortality Investigation (2004): *"if the explanatory variables themselves are as difficult to predict as the dependent variables (or indeed more so), then the projection's reliability will not be improved by including them in the model"*. Beyond this, Wilmoth (1998) observed that *"even if we understood these interactions and wanted to predict future mortality on the basis of a theoretical model, we would still need to anticipate trends in each of its components"* and, hence, we are still left with a problem of extrapolation.

We, therefore, believe that, whilst the analysis of the exogenous causes of changing mortality rates is important to understanding the past, it is not a useful method for making long-term projections into the future. We are forced by necessity to adopt an extrapolative approach for most practical purposes, especially those requiring model-based stochastic forecasts of mortality rates, such as risk management in the life insurance industry.

Furthermore, exponents of exogenous cause based models often start from the assumption that we exist at a privileged point in human history. This assumption can be optimistic, as in de Grey (2006), which argued that revolutions in the understanding of human genetics and biology just around the corner. Alternatively, this assumption can be pessimistic, as in Olshansky et al. (1998), which argued that we are approaching a hard limit in human longevity based on the fundamental obsolescence programmed into the human body beyond reproductive ages, or in Olshansky et al. (2005), which argued that the rise in obesity will soon threaten the increases in longevity we have witnessed to date. However, arguments such as these are not unique to the present time. The past century and more has been one of continuous medical progress (antibiotics, vaccinations, transplants, etc), but with new threats continuously arising (smoking, HIV/AIDS, obesity, etc). Wilmoth (1998) pointed out that *" extrapolations of past mortality trends assume, implicitly, a continuation of social and technological advance on a par with these earlier achievements"*. However, at every point within the past century, there were individuals making arguments very similar to those seen today, and that the time in which they lived was unlike any other which had come before and would come afterwards. What has been observed, however, is a steady improvement in human health and longevity, which is remarkable both for its endurance and its regularity. It is *"This combination of stability and complexity should discourage us from believing that singular interventions or barriers will substantially alter the course of mortality decline in the future"* (Wilmoth (1998)).

We, therefore, dispute the argument that consistency between the past and future when projecting mortality rates is, in fact, a weakness of the extrapolative approach. If we wish to understand what changing mortality rates look like during periods of rapid changes in medicine, lifestyle and society, then that information is available in the historical record. In analysing UK mortality data since 1950, for instance, we are basing our forecasts on a period of time which has witnessed far-reaching changes in lifestyle (for instance, the prevalence of smoking and the impact of diet on health) and medicine. It is also a period which saw a number of influenza pandemics (in 1951, 1957/58, 1968/69 and 2009) as well as the emergence of new diseases such as HIV. In short, we believe that careful analysis of the past and projections based upon this analysis if fully able to accommodate these

criticisms of the extrapolative approach.

When making extrapolative forecasts of mortality, it is, therefore, important to construct a mortality model which is fully capable of capturing the information in the historical data. For that reason, in Section 6.3, we use the "general procedure" (GP) described in Chapter 5 to construct a mortality model which can identify as much of the structure in the historical data as parsimoniously as possible. However, in order to make projections, we must ensure that the time series processes used to project the parameters in such models can replicate the features observed in the past. To this end, in Section 6.4, we introduce a method for detecting and projecting trend changes in the period parameters, and so address, at least partially, the criticism in Booth (2006) quoted above. For the cohort parameters, however, we suffer from the issue that we only have incomplete observations on generations which are still alive, and therefore require that the uncertainty in our parameter estimates for currently living generations blends smoothly into our projections for future years of birth. We discuss how this can be achieved in Section 6.5.

This is not to say, however, that events unprecedented in the historical record could not occur in future and have an important impact on future mortality rates. However, by definition, such events cannot be anticipated in advance and all attempts to do so are, necessarily, somewhat spurious. Certainly, unprecedented events should not form the basis of a "best estimate" of future mortality rates, but should only be included as unusual or "extreme" scenarios. Exploring the impact of unprecedented events via scenario analysis can be a useful tool to explore some extreme situations. However, it cannot perform any degree of quantification of the risk of these events occurring. In addition, the extrapolative approach is useful for establishing where such "extreme" scenarios should start from, by defining the limits of what is normal. For example, an extreme scenario based on an event unprecedented in the historical record must, by definition, produce an impact that is greater than, say, two standard deviations of the central forecast produced by an extrapolative approach using the past 50 years of data. However, we believe that such a subjective scenario analysis should only be performed after statistical and extrapolative projections have been produced, to examine the reasonableness of the projections, give insights into the tails of the projected distribution of mortality rates and allow results to be communicated with non-specialist stakeholders, rather than as the primary means of forecasting the future.

6.3 Fitting the past and identifying the model

We first use the GP to construct a suitable mortality model for data from the Human Mortality Database (2014) for men aged 0 to 100 in the UK over the period 1950 to 2009. The GP constructs a bespoke mortality model in the class of age/period/cohort models discussed in Chapter 2, of the form

$$\ln(\mu_{x,t}) = \alpha_x + \sum_{i=1}^{7} f^{(i)}(x;\theta^{(i)})\kappa_t^{(i)} + \gamma_{t-x} \tag{6.1}$$

where

- age, x, is in the range $[0, 100]$, period, t, is in the range $[1950, 2009]$ and therefore that year of birth, y, is in the range $[1850, 2009]$;
- α_x is a static function of age;
- $\kappa_t^{(i)}$ are period functions governing the evolution of mortality with time;
- $f^{(i)}(x;\theta^{(i)})$ are parametric age functions (in the sense of having a specific functional form selected a priori) modulating the impact of the period function dynamics over the age range, potentially with free parameters $\theta^{(i)}$;[1] and
- γ_y is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as it ages.

A summary of the terms in the models and their demographic significance[2] is given in Table 6.1 and the age and period functions shown in Figures 6.1a and 6.1b, respectively.

Many mortality models are not fully identified. This means that we can find transformations of the parameters[3] in the model which leave the fitted mortality rates unchanged. To uniquely specify the parameters, we impose identifiability constraints. These constraints are arbitrary, in the sense that they do not affect the fit to data, but they do allow us to impose our desired demographic significance on the terms in the model. These issues are discussed in detail in Chapters 3 and 4.

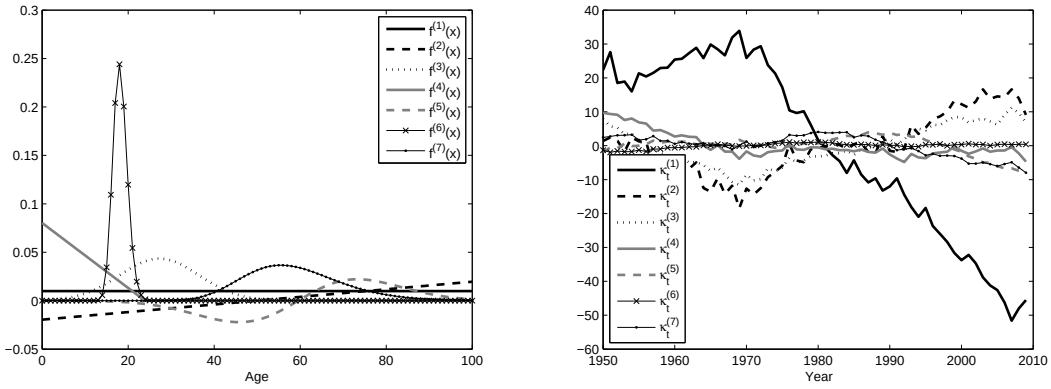
[1] For simplicity, the dependence of the age functions on $\theta^{(i)}$ is supressed in the notation used in the remainder of this chapter, but not in the model itself.

[2] Demographic significance is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

[3] These are called "invariant transformations" in Chapters 3 and 4.

Term	Description	Demographic significance
α_x	Static age function	Constant shape of mortality curve
$f^{(1)}(x)\kappa_t^{(1)}$	Constant age function	Level of mortality curve
$f^{(2)}(x)\kappa_t^{(2)}$	Linear age function	Slope of mortality curve
$f^{(3)}(x)\kappa_t^{(3)}$	Gaussian age function	Young adult mortality
$f^{(4)}(x)\kappa_t^{(4)}$	"Put option" age function	Childhood mortality
$f^{(5)}(x)\kappa_t^{(5)}$	Rayleigh age function	Postponement of old age mortality
$f^{(6)}(x)\kappa_t^{(6)}$	Log-normal age function	Peak of accident hump
$f^{(7)}(x)\kappa_t^{(7)}$	Gaussian age function	Late middle / old age mortality
γ_y	Cohort parameters	Lifelong year of birth effects

TABLE 6.1: Terms in the final model of Chapter 5



(A) Age functions (B) Period functions

FIGURE 6.1: Age and period functions for the mortality model

In the context of the model generated by the GP for the UK, we impose the following standard identifiability constraints

$$\sum_t \kappa_t^{(i)} = 0 \; \forall i \tag{6.2}$$

$$\sum_x |f^{(i)}(x)| = 1 \; \forall i \tag{6.3}$$

These identifiability constraints, respectively, allow us to:

- set a consistent level for each of the period functions, so that they represent deviations from an "average" level of mortality in the period, and
- select age functions a priori so that they have a consistent normalisation scheme. This enables us to compare the magnitudes of the period functions with each other and between populations and gauge their relative importance.

However, using the results of Chapter 4, we observe that the following transformations involving the cohort parameters leave the fitted mortality results unchanged[4]

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\alpha_x - a_0, \kappa_t^{(1)}, \kappa_t^{(2)}, \gamma_y + a_0\} \tag{6.4}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{\alpha_x + a_1(x - \bar{x}), \kappa_t^{(1)} - a_1(t - \bar{t}), \kappa_t^{(2)}, \gamma_y + a_1(y - \bar{y})\} \tag{6.5}$$

$$\begin{aligned}\{\hat{\alpha}_x, \hat{\kappa}_t^{(1)}, \hat{\kappa}_t^{(2)}, \hat{\gamma}_y\} = \{&\alpha_x - a_2((x - \bar{x})^2 - \sigma_y + \sigma_t), \kappa^{(1)} - a_2((t - \bar{t})^2 - \sigma_t), \\ &\kappa_t^{(2)} + 2a_2(t - \bar{t}), \gamma_y + a_2((y - \bar{y})^2 - \sigma_y)\}\end{aligned} \tag{6.6}$$

The degrees of freedom represented by the free parameters a_0, a_1 and a_2 in these transformations need to be used to impose three identifiability constraints on the cohort parameters when fitting the model. We choose these to be

$$\sum_y n_y \gamma_y = 0 \tag{6.7}$$

$$\sum_y n_y \gamma_y (y - \bar{y}) = 0 \tag{6.8}$$

$$\sum_y n_y \gamma_y ((y - \bar{y})^2 - \sigma_y) = 0 \tag{6.9}$$

where n_y is the number of observations of each cohort in the data. The justification for these constraints is that they appear to remove polynomial trends up to quadratic order in the cohort parameters at the fitting stage, so that they conform better with the demographic significance described in Chapter 2, i.e., that the cohort parameters should be centred around zero and not have any long-term trends. It is important to note that the choice of these constraints is still arbitrary and it is important that they do not affect our projections of mortality rates. This will influence our choices for the time series models we use to project the parameters in Sections 6.4.1 and 6.5 below.

6.4 Period functions

The fitted period parameters given in Figure 6.1b exhibit the following features:

[4] Here, $\bar{x} = \frac{1}{X}\sum_x x$, $\sigma_x = \frac{1}{X}\sum_x (x - \bar{x})^2$ and X is the number of ages in the data, and similarly for $\bar{t}$, $\bar{y}$, etc. These constants have been introduced to maintain the constraint that $\sum_t \kappa_t^{(i)} = 0$. Also note that, to aid understanding these complex relationships, Equations 6.4, 6.5 and 6.6 do not incorporate the normalisation factors required on the age functions in order to ensure that $\sum_x |f^{(i)}(x)| = 1 \; \forall i$. These will need to be included before the model is fitted to data.

- Most (but not all) of them appear to be non-stationary, according to statistical tests such as the ADF test.[5]
- The time series appear to be correlated, sometimes highly so. For instance, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ have a sample correlation of 92.6%.
- Some of the time series appear to show one or more changes in trend over the period.[6]

Our projections of the parameters should incorporate these features to ensure that our forecasts of the future are not systematically different from the structure observed in the historical data.

The period functions in mortality models have typically been projected using random walks with drift

$$\kappa_t^{(i)} = \kappa_{t-1}^{(i)} + \mu_0^{(i)} + \epsilon_t^{(i)} \tag{6.10}$$

The use of this process for the period functions runs from the earliest stochastic mortality model in Lee and Carter (1992), through Cairns et al. (2006a), to the more recent models in Plat (2009a), Cairns et al. (2011a) and Haberman and Renshaw (2011). In some cases, this time series process was selected after performing a Box-Jenkins analysis (e.g., Lee and Carter (1992)). In others, the process was chosen a priori without any statistical justification (e.g., Cairns et al. (2006a)), but based on its ability to produce biologically reasonable[7] forecasts of mortality rates.

The random walk with drift model has a number of desirable characteristics which make it an attractive process to use when projecting the period functions. It has a definite trend, allowing for mortality rates to decrease with time. It also has non-stationary variation around this trend, i.e., our projections get more variable as we make forecasts further into the future, which is important for making long-term projections. Further, it is not mean-reverting around this trend, and has a long memory of historical mortality shocks. By giving non-stationary and correlated period functions, the multivariate

[5] In particular, $\kappa_t^{(6)}$ is found to be stationary at the 5% level, whilst all other period functions are found to be non-stationary.

[6] Some studies refer to these as "structural breaks" rather than "trend changes". In this study, we use the terms "trend change" and "structural break" as synonyms.

[7] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

random walk with drift is also consistent with the first two observations in the historical data described above.[8] For these reasons, a multivariate random walk with drift has become the standard process for projecting period functions in mortality models and has therefore been used in the "naïve" projections of mortality we introduce in Section 6.6.

6.4.1 Identifiability of projections

When projecting mortality rates, it is important to use time series processes which do not depend on the arbitrary identifiability constraints we imposed on the parameters when fitting the model to data. Since these choices did not affect our analysis of the past, it is important that they do not affect our projection of the future. In particular, we must be certain that any conclusions we draw when using these models do not depend on earlier arbitrary choices when fitting the model.[9] We call time series which give projected mortality rates that are independent of the identifiability constraints "well-identified".

How to obtain well-identified projection methods was discussed in depth in Chapters 3 and 4 in the context of general APC mortality models. Chapter 3 established that the period functions from age/period mortality models should be projecting using time series processes which

- are multivariate, to allow for any potential correlations between the time series, and
- do not treat the different period functions differently. In practice, this means that the various time series should be integrated to the same order (in this case, $I(1)$).

This gives us an initial set of requirements, which are satisfied by using a multivariate random walk with drift process to project all of the period functions in the model.

However, the analysis of Chapter 4 discussed the more complicated identifiability issues present in models with a cohort term. These were caused by the collinearity between age, period and year of birth and meant that, in some models, specific deterministic trends

[8] However, in order to incorporate changes in trend, we must go beyond the random walk with drift process, which we do in Sections 6.4.2 and 6.4.3 below.

[9] In addition, identifiability constraints that are sensible when estimating the model might not be the most suitable when making projections. We will see this in Section 6.5 where we choose to change the identifiability constraints from those imposed when fitting the model to a new set which is more helpful when projecting the cohort parameters. We therefore need to ensure that our results are not affected by our new choice of identifiability constraints.

were unidentifiable, i.e., they could not be allocated between the age/period terms and the cohort term by the model and so required additional identifiability constraints to make this allocation manually. The allocation of these unidentifiable trends was performed using the transformations in Equations 6.4, 6.5 and 6.6, which were used to obtain a set of parameters satisfying the arbitrary constraints.

In this study, we apply the analysis of Chapter 4 to extend the random walk with drift process, as discussed in Section 6.4.1.1 below. Doing so, we ensure that the same time series process is appropriate for all possible sets of potential identifiability constraints and so will give the same projected mortality rates. However, in order to achieve this, there is a potential conflict between the second requirement from Chapter 3, namely that all period functions should be projected using the same processes, and the need to obtain biologically reasonable projections of mortality rates. This is discussed in Section 6.4.1.2, along with a potential resolution that provides projected mortality rates which are biologically reasonable and preserves the spirit of the requirements in Chapter 3.

6.4.1.1 First period function

Since the random walk with drift process is the most common time series process used to project the period parameters, we first need to show that it does not give well-identified projections of mortality rates for the model described in Section 6.3. We do this by showing that, if the random walk with drift process is suitable for $\kappa_t^{(1)}$ under one set of identifiability constraints, it will not necessarily be appropriate for a transformed $\hat{\kappa}_t^{(1)}$ under an alternative set of identifiability constraints.

To do this, first we note that Equation 6.5 adds a term linear in time to $\kappa_t^{(1)}$ and Equation 6.6 adds a term quadratic in time to $\kappa_t^{(1)}$. In addition, as discussed in Chapter 3, the level of $\kappa_t^{(1)}$ is undefined, meaning we can add a constant to it without changing the fitted mortality rates. Combining these, for $\kappa_t^{(1)}$, we write

$$\hat{\kappa}_t^{(1)} = \kappa_t^{(1)} + a_0^{(1)} + a_1^{(1)} t + a_2^{(1)} t^2 \tag{6.11}$$

This transformation converts one set of fitted parameters, satisfying one set of identifiability constraints, into an alternative set of parameters which satisfy a different set of identifiability constraints. These two sets of parameters, $\kappa_t^{(1)}$ and $\hat{\kappa}_t^{(1)}$ are equivalent: they give the same fitted mortality rates and so there is no statistical reason for preferring one over the other. As discussed in Chapters 3 and 4, this further implies that the

same time series process should be equally appropriate for either set of parameters.

For the time series process used for $\kappa_t^{(1)}$ to be appropriate for all equivalent sets of parameters (such as $\hat{\kappa}_t^{(1)}$), we need to make sure that it does not change form if we use the transformation in Equation 6.11 to move between $\kappa_t^{(1)}$ and $\hat{\kappa}_t^{(1)}$, i.e., that if $\kappa_t^{(1)}$ follows a random walk with drift process in Equation 6.10, then $\hat{\kappa}_t^{(1)}$ also follows a random walk with drift process. However, the random walk with drift process changes form when we apply the transformation in Equation 6.11 to it and, so, does not satisfy this requirement. We can see this by substituting the $\hat{\kappa}_t^{(1)}$ into the random walk with drift process to give

$$\begin{aligned}
\kappa_t^{(1)} &= \kappa_{t-1}^{(1)} + \mu^{(1)} + \epsilon_t^{(1)} \\
\hat{\kappa}_t^{(1)} - a_0^{(1)} - a_1^{(1)}t - a_2^{(1)}t^2 &= \hat{\kappa}_{t-1}^{(1)} - a_0^{(1)} - a_1^{(1)}(t-1) - a_2^{(1)}(t-1)^2 + \mu^{(1)} + \epsilon_t^{(1)} \\
\hat{\kappa}_t^{(1)} &= \hat{\kappa}_{t-1}^{(1)} + \mu^{(1)} + a_1^{(1)} - a_2^{(1)} + 2a_2^{(1)}t + \epsilon_t^{(1)}
\end{aligned}$$

We see that, if a random walk with drift was appropriate for $\kappa_t^{(1)}$, then a random walk where the drift changes linearly with time is appropriate for $\hat{\kappa}_t^{(1)}$. Hence, the random walk with constant drift is not appropriate for all equivalent sets of parameters and, therefore, all sets of identifiability constraints. This means that projections using such a process, the most commonly used in the literature to date, are not well-identified under the transformations in Equation 6.11 and, therefore, that we should not use it to project the model in Section 6.3.

However, this can be easily rectified. A random walk with drift process is not well-identified under the transformation in Equation 6.11 because the transformation introduced a term linear in time into the drift which was not present in the original time series. It is therefore natural to extend the random walk with drift process to introduce a term linear in time into the original time series. The transformation would then not add anything new to the process, merely modify what was already present. This suggests that we should use a random walk with linear drift for $\kappa_t^{(1)}$

$$\kappa_t^{(1)} = \kappa_{t-1}^{(1)} + \mu_0^{(1)} + \mu_1^{(1)}t + \epsilon_t^{(1)} \tag{6.12}$$

Again, we can check that this is well-identified by substituting $\hat{\kappa}_t^{(1)}$ into Equation 6.12 to confirm that we have the same time series process for both sets of parameters

$$\begin{aligned}
\kappa_t^{(1)} &= \kappa_{t-1}^{(1)} + \mu_0^{(1)} + \mu_1^{(1)} t + \epsilon_t^{(1)} \\
\hat{\kappa}_t^{(1)} - a_0^{(1)} - a_1^{(1)} t - a_2^{(1)} t^2 &= \hat{\kappa}_{t-1}^{(1)} - a_0^{(1)} - a_1^{(1)}(t-1) - a_2^{(1)}(t-1)^2 + \mu_0^{(1)} + \mu_1^{(1)} t + \epsilon_t^{(1)} \\
\hat{\kappa}_t^{(1)} &= \hat{\kappa}_{t-1}^{(1)} + \mu_0^{(1)} + \mu_1^{(1)} t + a_1^{(1)} - a_2^{(1)} + 2a_2^{(1)} t + \epsilon_t^{(1)} \\
&= \hat{\kappa}_{t-1}^{(1)} + \hat{\mu}_0^{(1)} + \hat{\mu}_1^{(1)} t + \epsilon_t^{(1)}
\end{aligned}$$

Although the numerical values we find for $\mu_0^{(1)}$ and $\mu_1^{(1)}$ are different for different sets of parameters (and, hence, identifiability constraints), the form of the time series is not. Hence, if a random walk with linear drift is appropriate for $\kappa_t^{(1)}$, it is also appropriate for $\hat{\kappa}_t^{(1)}$, and so, in turn, it is appropriate for all different sets of identifiability constraints. Therefore, the random walk with linear drift is well-identified.

We may find that under some sets of identifiability constraints, $\mu_1^{(1)}$ takes an apparently low value, and so we might be tempted to ignore it. Alternatively, we might be tempted to fit a random walk with linear drift and then test $\mu_0^{(1)}$ for statistical significance, with a view to setting it to zero. However, as shown above, the magnitude of $\mu_1^{(1)}$ is entirely dependent upon the identifiability constraints used, i.e., even if $\mu_0^{(1)}$ is close to zero, $\hat{\mu}_0^{(1)} = \mu_0^{(1)} + 2a_2$ can be arbitrarily large depending upon the value of a_2. Therefore any decision to ignore $\mu_1^{(1)}$ would also be entirely dependent upon the arbitrary identifiability constraints. Thus, the choice of time series to use for $\kappa_t^{(1)}$ cannot be motivated by arguments based on statistical significance or goodness of fit, but must be determined by the identifiability issues present in the model. Hence, we must use a random walk with linear drift for $\kappa_t^{(1)}$, regardless of the apparent size of $\mu_0^{(1)}$ to avoid generating poorly-identified projections of mortality rates that depend on the arbitrary constraints imposed when fitting the model.

In summary, the transformation in Equation 6.6 means that we must allow for quadratic trends in the first period function in the model. We do this by extending the conventional random walk with drift model to a random walk with linear drift process. This time series process is not changed fundamentally by changing from one set of identifiability constraints to another, and therefore will give projections which do not depend on the specific set of identifiability constraints adopted.

6.4.1.2 Other period functions

As with $\kappa_t^{(1)}$, we find that the invariant transformation in Equation 6.6, plus the unidentifiable level in the period functions, means that the fitted mortality rates are unchanged by a transformation of $\kappa_t^{(2)}$ in the form of

$$\hat{\kappa}_t^{(2)} = \kappa_t^{(2)} + a_0^{(2)} + a_1^{(2)} t \tag{6.13}$$

and of the form

$$\hat{\kappa}_t^{(i)} = \kappa_t^{(i)} + a_0^{(i)} \quad i = 3, \ldots 7 \tag{6.14}$$

for the other period functions.

A similar analysis to that performed in Section 6.4.1.1 shows that both the random walk with constant drift and the random walk with linear drift processes are well-identified under these transformations. The conclusions of Chapter 3, described at the beginning of this section, suggest that we should use random walks with linear drifts for all seven period functions in the model in order avoid treating $\kappa_t^{(1)}$ differently from the other period functions. However, if we do so, however, we obtain projections which are not biologically reasonable.[10]

We therefore have a conflict between our desire for projections which are biologically reasonable, on the one hand, and well-identified, on the other. Such conflicts were discussed in Chapter 3, where it was concluded that it was possible to treat age/period terms with parametric age functions as distinct, since there was no invariant transformation of the model which forced them to be interchangeable. However, using the same processes to project all the period functions in a model was still highly desirable because it was unlikely that the demographic significance of the term would lead to specific requirements for how it should be projected.

Models produced by the GP have parametric age functions, where each age function has a defined functional form, selected in advance of fitting the model to data to give each term distinct demographic significance. We, therefore, feel that it is justifiable to preserve this distinctiveness in order to ensure that the projections of mortality rates from

[10] Experiments have shown that using random walks with linear drifts for all period functions produces projections where mortality rates at some ages are predicted to continue decreasing and then start increasing in the near future with probability close to unity under this model. Without any biological reason why this should be the case, we consider this model to be inconsistent with existing medical knowledge.

the model are biologically reasonable. Furthermore, it is only the drift term which varies between the random walk processes used for the different period functions. We do not assume that the period functions differ in terms of stationarity, dependence structure or any other statistical property. Therefore, we feel that the use of a random walk with linear drift for $\kappa_t^{(1)}$, but random walks with constant drifts for the other period functions minimises the extent to which the various period functions are treated differently. This compromise preserves the spirit of the requirements in Chapter 3, whilst maintaining biologically reasonable projected mortality rates.[11]

Accordingly, we will use the random walk with drift model for the second to seventh period functions, but must use a random walk with linear drift process for $\kappa_t^{(1)}$ for the identifiability reasons discussed in Section 6.4.1.1, i.e., we use

$$\kappa_t^{(i)} = \begin{cases} \kappa_{t-1}^{(i)} + \mu_0^{(i)} + \epsilon_t^{(i)} & \text{if } i \neq 1 \\ \kappa_{t-1}^{(i)} + \mu_0^{(i)} + \mu_1^{(i)} t + \epsilon_t^{(i)} & \text{if } i = 1 \end{cases} \tag{6.15}$$

with innovations, $\epsilon_t^{(i)}$, which are allowed to be contemporaneously correlated.

As stated at the start of Section 6.4, we observed changes in trend in the historical period functions. However, the random walk model is not capable to reproducing this trend changes in future, even when it is well-identified. Consequently, we extend the random walk with drift model to allow for changes in trend, as described below.

6.4.2 Historical trend changes

We observed in Section 6.3 that some of the period functions appear to exhibit sharp changes in trend, which should be allowed for when projecting the model. A number of other studies have sought to detect and analyse changes in trend in mortality models using econometric techniques to detect structural breaks, for instance, Coelho and Nunes (2011), Sweeting (2011), Börger and Ruß (2012) and O'Hare and Li (2012b) which we discuss below.

Another, conceptually similar approach is to use "regime change" models, such as in Milidonis et al. (2011), Hainaut (2012) and Lemoine (2014). All of these studies have

[11] However, we note that it is theoretically possible to construct mortality models which give exactly the same fitted mortality rates to the model presented in Section 6.3, but which would require different time series processes to give well-identified projections and, hence, may give different projected mortality rates.

the disadvantage, however, that they assume only a finite number of regimes (usually two) and, therefore, discount the possibility for more radical changes in the evolution of mortality in future.

Our approach is to follow the structural break literature and accommodate changes in trend by allowing the drift functions for the random walks to be subject to infrequent and random jumps, i.e., we replace

$$\mu_0^{(i)} \text{ with } \mu_0^{(i)} + \sum_{j=1}^{N^{(i)}} \nu_j^{(i)} I_{t \geq \tau_j^{(i)}} \quad i \neq 1 \quad \text{and}$$
$$\mu_1^{(1)} t \text{ with } \mu_1^{(1)} t + \sum_{j=1}^{N^{(1)}} \nu_j^{(1)} (t - \tau_j^{(1)})^+ \tag{6.16}$$

in the random walk model in Equation 6.15, where $N^{(i)}$ is a Poisson counting process for the number of changes in trend occurring at times $\tau_j^{(i)}$, $j = 1, \ldots, N^{(i)}$,[12] I is an indicator value and $x^+ = max(x, 0)$.

To allow for changes in trend in future, we must first identify the trend changes that are present in the historical data. This, in part, addresses the criticism of Booth (2006) raised in Section 6.2. A number of methods have been proposed to do this.[13] We use the method developed in Bai and Perron (1998),[14] since it is capable of identifying multiple structural breaks and we find it to be relatively intuitive to implement. An outline of this procedure is given below, but it is discussed in greater detail in van Berkum et al. (2014):

- Each period function is considered independently.
- Conditional on k trend changes occurring at dates $\tau_j^{(i)}$ in period function i, the magnitude and direction of the trend changes $\nu_j^{(i)}$ can be calculated using least squares regression, as well as the log-likelihood and Bayes Information Criterion (BIC)[15] of the observed time series.
- Conditional on k trend changes occurring, we test every possible set of dates for the trend changes to select the values of $\tau_j^{(i)}$ which maximises the log-likelihood

[12] By convention, $\sum_{j=1}^{N^{(i)}} X_j = 0$ for $N^{(i)} = 0$.

[13] For instance, Sweeting (2011) and Börger and Ruß (2012) use a method based on the DW (Durbin and Watson (1951)) statistic to identify multiple trend changes in a trend-stationary process and Coelho and Nunes (2011) use the method of Harris et al. (2009) in conjunction with testing for a unit root.

[14] This technique is also used in O'Hare and Li (2012b) and van Berkum et al. (2014).

[15] Defined as max(Log-likelihood) $- 0.5 \times$ No. free parameters $\times \ln$(No. data points).

(and BIC) of the fitted time series. Consistent with van Berkum et al. (2014), to prevent over-fitting the model and finding spurious changes in trend, we assume that trend changes cannot occur within five years of each other, or within the first and last five years of the dataset.

- k is then increased sequentially until the BIC has stopped increasing to give $N^{(i)} = argmax$ BIC(k).

It is important to note that this procedure can be very computationally intensive if long datasets are used, and so may cause practical issues in implementation. Bai and Perron (2003) presented an approach for detecting multiple structural breaks in time series which gives the same results as the procedure described above, but is based on dynamic programming and is considerably faster to implement. Since we only consider 60 years of data in this study, this techniques was not used in this study. However, we acknowledge that, in order to gain a more comprehensive understanding of the dynamics of trend changes, a longer period of data is required, as in Sweeting (2011) and Börger and Ruß (2012).[16]

i	$\mu_0^{(i)}$ Constant drift	$\mu_1^{(i)}$ Linear drift	$N^{(i)}$ No. trend changes	$\tau_j^{(i)}$ Date of trend change	$\nu_j^{(i)}$ Size of trend change
1	0.0350	-0.0397	0	N/A	
2	0.1036	N/A	0	N/A	
3	-1.0236	N/A	1	1970	1.4842
4	-0.2441	N/A	0	N/A	
5	0.1296	N/A	1	1993	-0.7905
6	0.0295	N/A	0	N/A	
7	-0.1766	N/A	0	N/A	

TABLE 6.2: Fitted time series parameters for the period functions

Using this procedure, we obtain the estimates of the time series parameters in Equations 6.15 and 6.16 given in Table 6.2 for the historical period functions (without allowing for parameter uncertainty). We have not attempted to relate the timing and direction of these trend changes to specific underlying socio-economic drivers of mortality for the population as such relationships would be highly speculative. The model detects just two significant trend changes in seven period functions, each over 60 years of data. This is a comparatively small number, which makes a sophisticated statistical analysis of the nature of the trend changes impossible. Accordingly, we must make a number of simplifying assumptions in order to project trend changes in the future.

[16] However, using longer periods of data runs into problems caused by the jumps in mortality rates during the First and Second World Wars. This is why we have limited our analysis to only use data since 1950.

6.4.3 Projecting trend changes

It is desirable that projections of future mortality are consistent with the features which have been observed in the historical data. Just as we have seen that the historical data contains structural breaks where the trend rate of improvement in mortality has changed, so we can envision scenarios where these may occur at an unknown point in future, caused by medical breakthroughs in the treatment of disease or socio-economic changes in the population, for example. Accordingly, we should project changes in the trends in our parameters to occur in future if they have been detected in the past. This is in contrast to the work of Coelho and Nunes (2011) and van Berkum et al. (2014), who do not allow for future trend changes in projections.

We believe that allowing for trend changes is also important for managing longevity risk, as an acceleration of the trend rate of improvements would dramatically increase the present value of annuity liabilities. To do so, we need to make assumptions on the dependence, frequency, direction and magnitude of potential trend changes in future, in order to give both biologically reasonable projections and to be as consistent with the observed historical trend changes as possible. Finally, we have a strong preference for simple, parsimonious models due to the small number of observed trend changes available to calibrate our models.

6.4.3.1 Dependence between period functions

We do not have sufficient observed data to be able to determine whether trend changes in the different time series are more or less likely to occur simultaneously. Specifically, we do not test for the phenomenon of co-breaking[17] and assume that breaks in the different time series occur independently of each other. This contrasts with the approach of Sweeting (2011), where trend changes were often observed simultaneously in the different period functions.

6.4.3.2 Frequency of trend changes

We assume that future trend changes occur with the same frequency as the historical trend changes observed in the fitted time series, e.g., if we observe two trend changes in a 60-year sample period for a period function, we assume that the probability of a trend change occurring in any projected year is $\frac{1}{30}$. We also assume that the number

[17] Defined in Hendry and Massmann (2005) as when structural breaks are observed in two or more time series, but not in a linear combination of them.

of trend changes is a Markov process and, accordingly, this probability does not change depending on when the previous trend change was observed. This is the same approach as was adopted in Sweeting (2011) and Börger and Ruß (2012).

This assumption may be considered to be unrealistic, since it could reasonably be argued that a trend change is more likely to be observed in a year if none have been observed for a long time. However, because only one trend change has been observed for any individual time series in the past, any more complex dependence structure would have to be justified in terms of the underlying biological and demographic processes driving the period functions. Since such a justification would, necessarily, be highly subjective, we opt for a Markov process for simplicity.

Nevertheless, we should be aware that this assumption has a number of weaknesses. First, the Markov assumption is inconsistent with the restriction that trend changes in the historical data cannot occur within five years of each other. Although the projection method can easily be modified to allow for a minimum length of time between trend changes, in practice, the probability of two projected trend changes occurring within five years is very low. When we restricted projected trend changes so they could not occur within five years of each other, it made little difference to the projections of mortality rates.

Second, it implies that for time series where no trend change has been observed in the past, we assume with certainty that no trend change can occur in future. This is unavoidable, since even if we were to allow for a non-zero chance of trend changes occurring in future in these time series, we would have no data to calibrate the magnitude of any changes. This problem can be mitigated to an extent by allowing for parameter uncertainty in the fitted period functions using a bootstrapping method such as that developed in Koissi et al. (2006). This generates a large number of pseudo-datasets by bootstrapping the fitted residuals from the original model to give resampled death counts. The model in Section 6.3 is refitted to each of these sets of death counts, giving a re-estimate of the different period functions and, hence, an estimate of the level of parameter uncertainty in them. These resampled estimates of the period functions are then tested individually for the number and timing of trend changes. Thus, parameter uncertainty is allowed for, both in the period functions and in the parameters of the time series processes assumed to generate them. Because of this, we may identify trend changes in some sets of bootstrapped period functions, even when we did not detect any

in the original period function. We use this technique to generate the results shown in Figure 6.3 and in Section 6.6.

6.4.3.3 Direction of trend changes

We assume that trend changes are as likely to be positive as negative, i.e., there is an equal chance of them improving mortality as worsening it. The limited number of historical trend changes means that any specific assumption on the direction of a trend change would need to be justified by the underlying socio-economic drivers of mortality. Biological and demographic arguments can be made on either side to support the case that the decreases in mortality rates currently observed will cease in future (for instance, the rise of obesity in the population, as discussed in Olshansky et al. (2005)) or that breakthroughs in medical progress will lead to an acceleration of the improvements in mortality (for instance, see de Grey (2006)). In light of this "great debate"[18] in demography, we remain agnostic as to whether future changes in trend are more likely to improve or worsen mortality rates at this point.

Our chosen model for projected trend changes leaves the median forecast of mortality unchanged (compared with a model which extrapolated the most recent observed trend), but affects the tails of the projected distribution. This is consistent with the notion that extrapolating the most recent past represents a "best estimate" of future improvements in mortality in the short run. Allowing for changes in trend, however, is important for risk management purposes, as discussed in Section 6.6.3.

6.4.3.4 Magnitude of trend changes

The magnitude of projected trend changes is the most subjective of the assumptions we need to make. Sweeting (2011) and Börger and Ruß (2012) assume that the magnitude of a trend change is normally distributed, with mean and standard deviation calibrated from the observed values. Instead, we assume that the magnitude of the trend changes follows a Pareto distribution, based on a consideration of the trend changes we have observed.

All methods of detecting trend changes in the historical data will fail to detect genuine but small trend changes, since these will not be found to be statistically significant.

[18] So named by Siegel (2005).

Consequently, the trend changes found in the past and available for analysis are not representative of the full distribution of trend changes, but merely a truncation of this distribution. Assuming that the threshold size for a trend change to be detected is sufficiently "large", the relevant distribution for the observed trend changes will therefore be the Pareto distribution, regardless of the "true" underlying distribution for the trend changes.

Our projections of mortality from the model should be consistent with what was observed in the past. A model that projects trend changes which it could not have found in the data violates this consistency. The Pareto distribution can generate future trend changes which are above the threshold for statistical significance and therefore will be consistent with those observed in the past. We believe that this is preferable to the methods used in Sweeting (2011) and Börger and Ruß (2012), which may generate a significant number of "small" trend changes which could not have been detected had they occurred in the historical data.

Another desirable property of the Pareto distribution is that it is long tailed and so can generate some very large future trend changes, which may be useful for the risk assessment of extreme mortality scenarios. It also has only two parameters for each period function - the size of the threshold, $\nu_{crit}^{(i)}$, and a scale parameter, $\alpha^{(i)}$ - which are relatively easy to estimate based on the limited number of historical observations.

The threshold, $\nu_{crit}^{(i)}$, for the period function can be approximated by considering the minimum size of a trend change that would be found to be statistically significant at a given confidence level.[19] Consider a period function generated by a random walk process with a trend change at $t = 0$, when the drift changed from known drift, μ, to $\mu + \nu$

$$\Delta\kappa_t = \begin{cases} \mu + \epsilon_t & \text{if } t \leq 0 \\ \mu + \nu + \epsilon_t & \text{if } t > 0 \end{cases}$$

where the magnitude of the change in drift, ν, is unknown. Considering the period $[1, T]$, where T is the average time between trend changes, we would obtain the least squares

[19] Although this is not the technique described in Section 6.4.2 above, it gives a threshold trend change size consistent with those seen in practice.

estimate

$$\begin{aligned}\hat{\mu} &= \frac{1}{T-1}\sum_{t=1}^{T-1}\Delta\kappa_t \\ &= \mu + \nu + \frac{1}{T-1}\sum_{t=1}^{T-1}\epsilon_t\end{aligned}$$

for the drift of the time series. If we were to perform a hypothesis test to determine whether the estimated drift in $[1,T]$, $\hat{\mu}$, is equal to the drift in the earlier period, μ, the null hypothesis would be that there was no change in trend. Therefore, we would reject the null hypothesis if

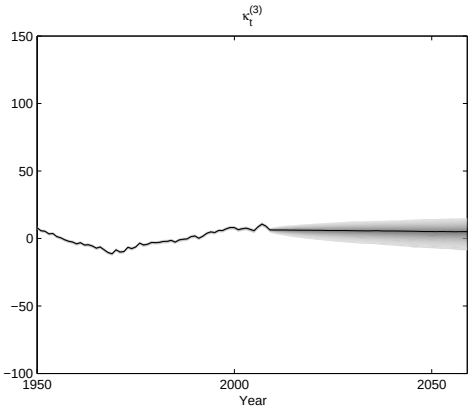
$$\begin{aligned}|\hat{\mu} - \mu| &= |\nu + \frac{1}{T-1}\sum_{t}\epsilon_t| \\ &\approx |\nu| \geq Z\frac{\sigma}{\sqrt{T-1}}\end{aligned}$$

i.e., we would only expect to detect trend changes above the threshold $|\nu| \geq Z\frac{\sigma}{\sqrt{T-1}}$, where Z is the critical statistic from the normal distribution at a given significance level and σ is the standard deviation of the innovations (which is assumed to be known but which can be estimated from our fitted period function). Similar considerations for the random walk with linear drift yield $|\nu| \geq Z\frac{6\sigma}{\sqrt{T(T-1)(2T-1)}}$. These values for each time series, with Z taken from the normal distribution at the 99% level, are then used as the threshold values $\nu^{(i)}_{crit}$ when generating Pareto random variables.

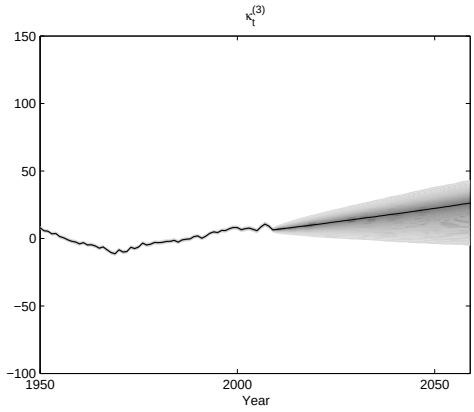
Once the threshold of the Pareto distribution has been estimated, the scale parameters, $\alpha^{(i)}$, can be estimated by matching the sample means of the observed trend changes, $\bar{\nu}^{(i)} = \frac{1}{N^{(i)}}\sum_{j=1}^{N^{(i)}}\nu^{(i)}_j$, with the mean of the theoretical distribution to give

$$\alpha^{(i)} = \frac{\bar{\nu}^{(i)}}{\bar{\nu}^{(i)} - \nu^{(i)}_{crit}}$$

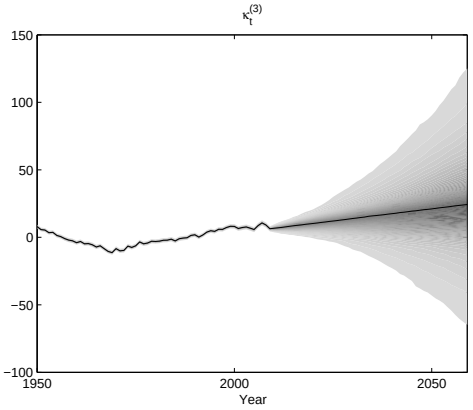
We derive values of $\nu^{(i)}_{crit}$ and $\alpha^{(i)}$ for each time series, $i = 1, \ldots 7$, and, when allowing for parameter uncertainty, for each set of resampled period functions. Thus, we also allow for parameter uncertainty in the distribution of future trend changes as well as allowing for uncertainty in their number and timing.



(A) Without trend changes



(B) With historical trend changes



(C) With historical and projected trend changes

FIGURE 6.2: 95% fan charts for projected period function, $\kappa_t^{(3)}$, under three different assumptions regarding trend changes

6.4.3.5 Impact of trend changes on projected period functions

Figure 6.2 shows fan charts of the 95% confidence intervals for the projected $\kappa_t^{(3)}$ period function using first a standard random walk with drift without allowing for historical or projected trend changes (Figure 6.2a), then allowing for historical trend changes but not projecting any in future (Figure 6.2b - similar to the approach in van Berkum et al. (2014)) and finally the approach discussed above (Figure 6.2c), allowing for both historical and projected trend changes. In all cases, parameter uncertainty is allowed for in the fitted parameters using the bootstrapping method discussed by Koissi et al. (2006).

It can be seen that allowing for trend changes alters the fan chart of the projected period function in a number of ways compared to the case when trend changes are not allowed.

- Allowing for trend changes gives different median projections for the period functions. The median projection from a random walk continues the trend found by

drawing a straight line between the first and last values of $\kappa_t^{(3)}$. By allowing for a trend change to occur during the historical period, our median projections extend the more recent trend operating since 1970, as shown in Figures 6.2a and 6.2b.

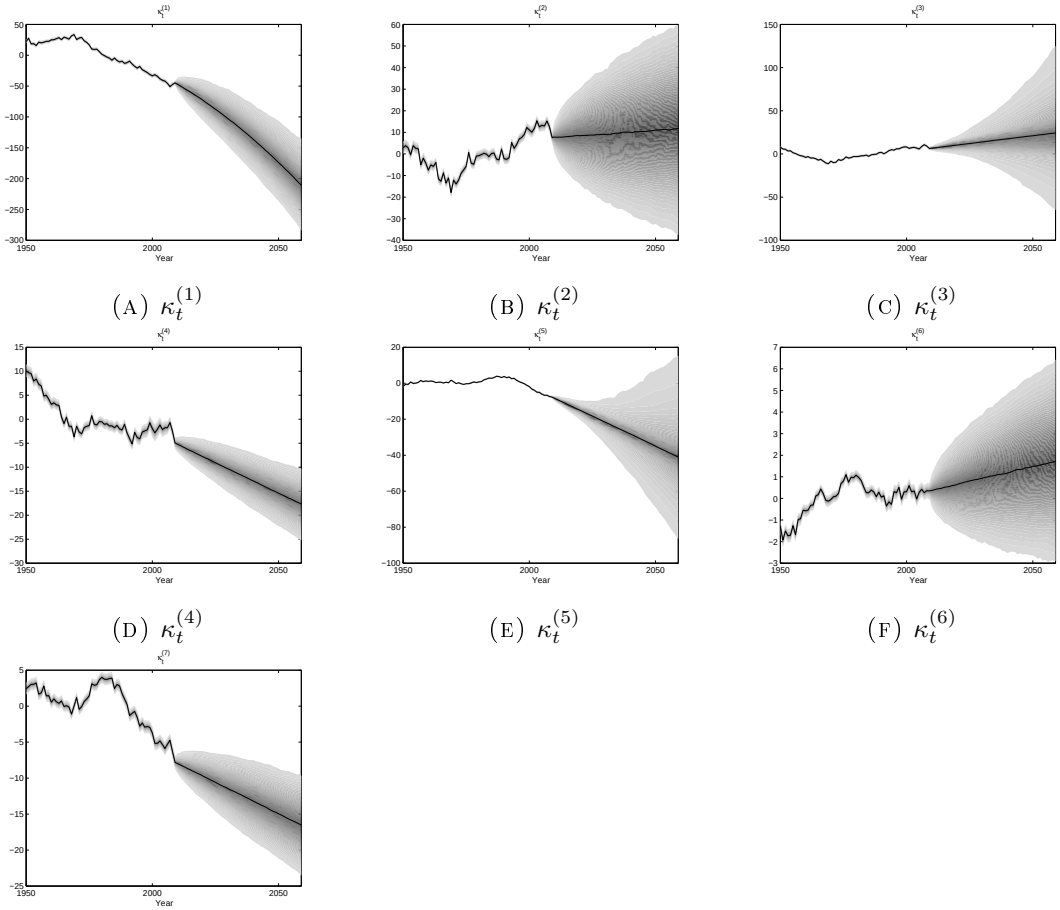
- Allowing for trend changes gives narrower projection intervals in the short run, as shown in Figure 6.2b. This is because our improved estimate of the trend in the historical period functions has reduced our measured variability around this trend. When this is projected, it leads to narrower projection intervals around the central trend in the short run (i.e., until we project a change in trend). We argue that this is more plausible as mortality rates in the near future are unlikely to be very different from a simple extrapolation of those observed today.

- Allowing for trend changes gives wider projection intervals in the long run, as shown in Figures 6.2b and 6.2c. This is because we allow for the central trend to change in future. Whilst the width of the projection interval from a random walk with drift will grow with projection time τ at the rate $\tau^{\frac{1}{2}}$, the projection interval from a random walk with a drift changing at random discrete intervals will grow at the rate $\tau^{\frac{3}{2}}$.[20] We argue that this is more plausible as the more distant future is highly uncertain, with numerous medical, demographic and socio-economic factors which might impact mortality rates radically in a fundamentally unpredictable manner.

All of these changes give projections which we consider to be more consistent with the historical period function, and allow for a more plausible assessment of the relative uncertainty of both the near and more distant future. This is despite these methods being fairly simple and yielding only quite crude estimates for the distribution of trend changes. However, we are constrained by the limited number of observed trend changes found in the historical data and therefore are prevented from using more sophisticated methods. We also feel that, since the purpose of our projections is to provide more plausible allowances for extreme longevity risk in projected mortality rates, any greater sophistication would be somewhat spurious. Fan charts for all of the projected period functions, allowing for parameter uncertainty using the residual bootstrapping method of Koissi et al. (2006), are shown in Figure 6.3.

In summary, we propose a method for detecting trend changes in the historical period functions, based on the approach in Bai and Perron (1998), and projecting future trend changes based on assumed distributions for the frequency, direction and magnitude of these future trend changes which are consistent with what has been observed in the past. We have taken steps to ensure that these time series are well-identified, in the

[20] A proof of this result, which is independent of the assumed distribution of the trend changes, is given in Appendix 6.A.



(A) $\kappa_t^{(1)}$ (B) $\kappa_t^{(2)}$ (C) $\kappa_t^{(3)}$

(D) $\kappa_t^{(4)}$ (E) $\kappa_t^{(5)}$ (F) $\kappa_t^{(6)}$

(G) $\kappa_t^{(7)}$

FIGURE 6.3: 95% fan charts for projected period functions with historical and projected trend changes

sense that our projections of mortality rates do not depend on the arbitrary identifiability constraints we imposed when fitting the model. Allowing for trend changes to occur in future gives projections which are considerably more uncertain, especially as we project further into the future, which we believe is more biologically reasonable and has significant impacts on risk management, as discussed in Section 6.6.3.

6.5 Cohort parameters

The cohort parameters in the model, shown in Figure 6.4, represent lifelong mortality effects specific to distinct years of birth which we interpret in terms of the life histories of the relevant cohorts in Chapter 5.

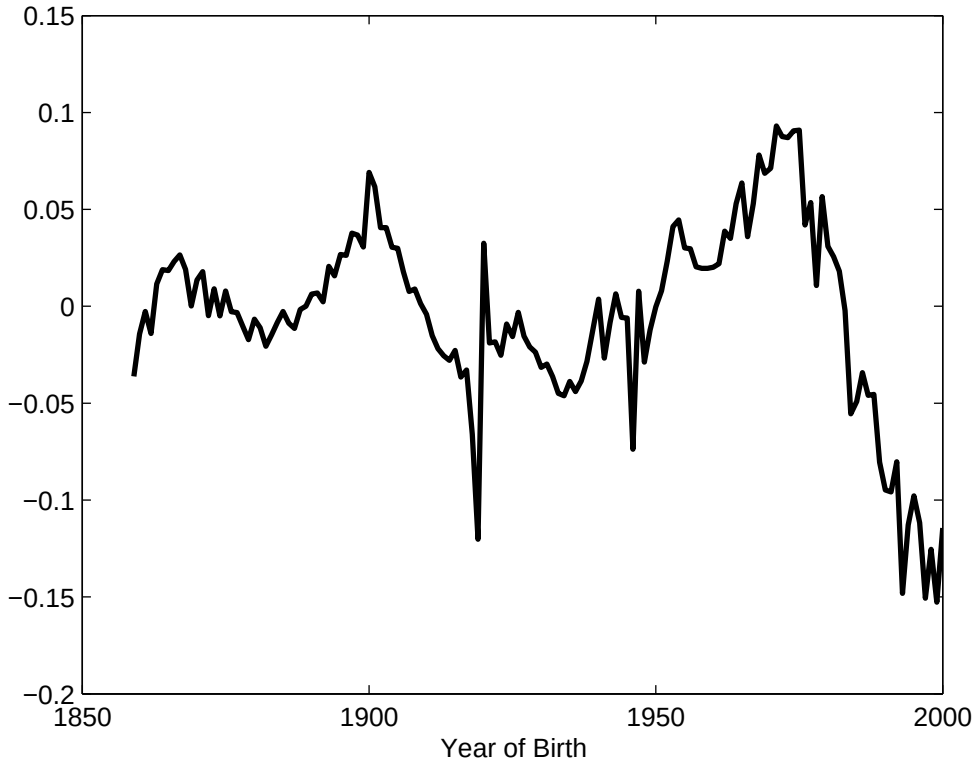


FIGURE 6.4: Cohort parameters

Given our desire for the cohort parameters to have the demographic significance discussed in Chapter 2, we would like our projections of the cohort parameters to have the following properties:

- The cohort parameters should represent genuine lifelong mortality effects, rather than merely being mis-classified age/period effects resulting from an incorrect specification of the model. This is an especially large problem for the most recent years of birth, since cohort parameters for these are only estimated on the basis of data at younger ages, where it is more difficult to properly specify the age/period terms in a model. We achieve this by using the general procedure to sequentially select age/period terms which capture all the significant age/period structure in the data, before adding a set of cohort parameters to the model.

- The cohort parameters should lack trends, i.e., have $\mathbb{E}\gamma_y = 0$ unconditionally for all y for both past and future years of birth. This is consistent with the notion that the cohort effects represent a deviation from the level of mortality for a "typical" cohort. We achieve this through careful choice of our identifiability constraints, as discussed in Section 6.5.2.

- The projected cohort parameters should be stationary, in the sense that the variability of the cohort parameters around the central trend should not change with time. We do not believe there is any compelling reason to suppose that the variability in the lifelong mortality factors should be any greater for future cohorts than for those observed to date. This is also consistent with the belief that cohort effects may persist for several years or decades, but should not result in permanent changes in the level of mortality, otherwise they should be re-classified as period effects.

- The projected cohort parameters should be independent of the period effects. We believe that cohort effects have very different demographic significance from the period effects and are treated separately when fitting the model. For a full discussion of this issue, see Chapter 4. In addition, an assumption of independence is both practical and parsimonious.

- The projection method used for the cohort parameters should take account of "unusual" birth cohorts, such as those in 1919/1920 and 1946/1947. Based on the analysis of Richards (2008) and Cairns et al. (2014), we believe that the unusual mortality rates associated with individuals born in these years are not due to genuine cohort effects, but are artefacts of the data. These are caused by the atypical and uneven pattern of births occurring in these years as a result of the demobilisations of soldiers after the First and Second World Wars, respectively, which, in turn, led to a mis-estimation of the size of the exposed population for those years of birth. A Third World War lies outside the scope of any mortality model to project, and therefore it seems reasonable not to allow for similar cohort effects to re-occur in future. Nevertheless, the observed cohort effects will persist in observed mortality rates in future. We accommodate this by allowing for indicator variables to capture the outliers in these years and deal with them in the historical parameters without affecting our estimates of the time series used to project the parameters into the future.

There is currently no well-established method for projecting the cohort parameters. A number of techniques are discussed in Cairns et al. (2011a) and van Berkum et al. (2014). Many of these fit time series from the ARIMA family in order to make projections. The classical approach to projecting the cohort function is to use Box-Jenkins methods to fit a preferred time series process to the historical cohort parameters and then to use this process to project them into the future. The limitations of this approach in obtaining projected parameters which have consistency between the past and future are discussed in Section 6.5.1. In addition, there is no guarantee that the preferred time series found by Box-Jenkins methods will be well-identified (i.e., they do not depend on

the identifiability constraints imposed in Section 6.3 when the model was fitted to data). We therefore discuss how well-identified cohort projections can be obtained in Section 6.5.2 and then offer a Bayesian approach which both gives well-identified projections and allows adequately for the uncertainty in the parameters in Section 6.5.3.

6.5.1 The classical time series approach

When fitting time series models to the cohort parameters, many authors use Box-Jenkins methods to select an appropriate model. Implicitly, these methods assume that the observed values of the time series are all known with the same degree of certainty. However, we have considerably less information about the latest cohorts than the earlier ones. It is therefore important to use methods which apply less weight to the later cohorts when estimating any time series parameters. Therefore, the classical Box-Jenkins framework is not appropriate.

To demonstrate this, consider the pattern of cohort effects shown in Figure 6.4 and, in particular, the most recent downward trend in the parameters dating from around 1975. Fitting a time series using standard Box-Jenkins methods would give these 25 years' worth of data points the same weight as the parameters covering the period from 1920 to 1945, for instance. However, the cohort effects for the most recent years of birth are considerably more uncertain for two reasons.

First, we have observed these cohorts for less time and so have fewer annual observations of them. For example, we have only 30 observations of the cohort born in 1980 in our data, whilst we have 90 observations of the cohort born in 1920. We recognised this was an issue in fitting the model to the extent that we did not attempt to estimate parameters for years of birth with fewer than ten observations. It would, therefore, be inconsistent to then disregard this issue when we come to project the cohort parameters.

Second, these cohorts comprise young people whom we would not expect to have died in large numbers during the period we have been observing them. Not only are we making estimates based on fewer observations, but these observations are associated with very few deaths. As a consequence, any conclusions on the mortality in these most recent cohorts is subject to very considerable uncertainty.

We can see this more formally by considering the Fisher information matrix under maximum likelihood estimation assuming the death count, $D_{x,t}$, for each age and period is a conditionally Poisson-distributed random variable, which will give a lower bound for the standard deviation of our parameter estimates via the Cramér-Rao bound:

$$\begin{aligned} I(\gamma_y) &= -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \gamma_y^2}\right] \\ &= \sum_x W_{x,y+x}\, E^c_{x,y+x}\, \mu_{x,y+x} \\ &= \sum_x W_{x,y+x}\, \mathbb{E}D_{x,y+x} \\ &\geq \frac{1}{\mathbb{V}ar(\gamma_y)} \end{aligned} \tag{6.17}$$

where $E^c_{x,t}$ are the observed central exposures to risk and $W_{x,t}$ are a set of weights for each age and period. This shows that the variance of a cohort parameter is inversely proportional to the number of deaths expected to date for that year of birth. The observed cohort parameters are therefore unavoidably heteroskedastic. In contrast, Box-Jenkins methods assume that the observations of the time series process under investigation are either known with certainty or estimated with the same degree of uncertainty, and so Equation 6.17 invalidates the traditional approach to selecting a time series model in these circumstances.

There are two potential "classical" methods which could be used to resolve this issue:

- We could fit an ARIMA time series process using a weighted least squares approach, and explicitly give less weight to cohort parameters felt to be more uncertain when estimating the time series parameters.
- We could allow for parameter uncertainty in our estimates of the historical cohort effects, for instance, by using Bayesian techniques (as in Pedroza (2006)) or by residual bootstrapping (as in Koissi et al. (2006)).

Both of these methods make some attempt to correct for the higher level of uncertainty in the recent cohort effects when we come to select a time series process and estimate the parameters within it.

However, classical approaches assume that the existing parameter estimates will not be revised in light of the new information that future data will contain. This, therefore, still assumes that there is a discontinuity between the "known" historical parameters

used to estimate the process and the unknown future parameters which are projected. This discontinuity leads to a sharp increase in the modelled level of uncertainty in the parameters between the historical parameters and the projected parameters.

While this is true for the period functions, since no new data obtained for future years will make us revise our estimate for $\kappa^{(1)}_{1975}$, it does not hold for the cohort parameters. This is because we will continue to observe cohorts born recently for decades into the future and use these observations to revise the estimated cohort parameters on an on-going basis. To illustrate, the last fitted cohort parameter we have is for year of birth 1999 and the first projected cohort parameter is for 2000. The classical approach would assume that γ_{1999} is known with certainty whilst γ_{2000} needs to projected. However, we will continue to observe both cohorts for nearly a century, and so our current estimate of γ_{1999} should be considered an approximation based on partial information and subject to future revision. In addition, we possess only slightly more information for estimating γ_{1999} than γ_{2000} and so the assumption that one is known whilst the other is unknown is inconsistent with the data we possess. In order to obtain a desired consistency between the historical and projected cohort effects, we use the Bayesian approach described in Section 6.5.3 which is capable of allowing for the incomplete nature of the information we have regarding cohorts which are currently alive when projecting the cohort parameters.

6.5.2 Identifiability in projections

In addition to the considerations discussed above, the use of Box-Jenkins methods to select a time series process for the cohort parameters can lead to the use of time series processes which are not well-identified. Just as in the discussion concerning identifiability in the period parameters in Section 6.4.1, we need to ensure that our projected mortality rates are well-identified, i.e., they do not depend on the identifiability constraints imposed. To change the identifiability constraints on the cohort parameters, we need to use the transformations in Equations 6.4, 6.5 and 6.6 to obtain a new (but equivalent) set of parameters. We therefore need to ensure that the time series process used for the cohort parameters does not change if we use these transformations and so are equally appropriate for all sets of identifiability constraints.

We see that Equation 6.4 adds a constant to γ_y, Equation 6.5 adds a term linear in year of birth to γ_y and Equation 6.6 adds a term quadratic in year of birth to γ_y. These can

be combined and written as

$$\hat{\gamma}_y = \gamma_y + a_0 + a_1 y + a_2 y^2 = \gamma_y + AX_y \tag{6.18}$$

where $X_y = \begin{pmatrix} 1, & y, & y^2 \end{pmatrix}^\top$. As with the period functions in Section 6.4.1, this transformation converts one set of fitted parameters (using one set of identifiability constraints) into an alternative set of parameters which satisfy a different set of identifiability constraints. These two sets of parameters, γ_y and $\hat{\gamma}_y$, are equivalent: they give the same fitted mortality rates and so there is no statistical reason for preferring one over the other.

As discussed in Chapter 4, identifiability under this transformation means that we need to allow for linear and quadratic trends within the cohort parameters, even if they are not apparent visually. The desire for a stationary distribution around these central, deterministic trends leads us to use an ARMA time series process of the form

$$\Phi(L)(\gamma_y - \beta X_y) = \Psi(L)\epsilon_y \tag{6.19}$$

where β is a matrix of regression coefficients found from analysing the fitted parameters and L is the lag operator. We can see that this is well-identified by applying the transformation in Equations 6.18 to Equation 6.19 to obtain an equivalent set of parameters, which we then substitute into Equation 6.19 to give

$$\Phi(L)(\hat{\gamma}_y - AX_y - \beta X_y) = \Phi(L)(\hat{\gamma}_y - \hat{\beta} X_y) = \Psi(L)\epsilon_y \tag{6.20}$$

Doing this has changed the numerical values of the regressors in β, but nothing fundamental about the time series, such as the moving average and autoregressive terms, Φ and Ψ. Hence, if the time series process was appropriate for γ_y, it is also appropriate for $\hat{\gamma}_y$ and, therefore, appropriate for all different sets of identifiability constraints. Hence, this time series model is well-identified.

The specific nature of the time series can be set by choosing the polynomials $\Phi(L)$ and $\Psi(L)$. In principle, these could be selected via a modified Box-Jenkins process, but taking care to include the βX_y term. Alternatively, we can work backwards from our desired demographic significance of the cohort parameters to select $\Phi(L)$ and $\Psi(L)$, whilst also including the βX_y term to ensure that the process is well-identified.

For instance, an AR(1) process, with $\Phi(L) = 1 - \rho L$ and $\Psi(L) = 1$, might be felt to be consistent with the desired demographic significance as it is stationary, parsimonious,

but still allows for persistent cohort effects. AR(1) processes are often used for the cohort parameters in mortality models, for instance in Cairns et al. (2011a). In order to make this well-identified, however, we could choose to project using an AR(1) process around a quadratic trend by including a βX_y term, as discussed above. This is the "AR(1) process around a quadratic drift" process discussed in Chapter 4 for the model of Plat (2009a).

When we project using the AR(1) process around a quadratic drift, we obtain $\mathbb{E}\gamma_y = \beta X_y$ unconditionally. Consequently, it might be felt that there is a conflict between the need for the time series process to be well-identified and our desired demographic significance for the cohort parameters, namely that they lack trends. We need to allow for quadratic trends in order to give well-identified projections, but we would like these trends to be zero based on our (subjective) demographic significance, i.e., we would like to have $\beta = 0$. Clearly, the need to have well-identified projections which do not depend upon arbitrary identifiability constraints is more important. However, it is possible to achieve both aims simultaneously.

As shown by Equation 6.20, the value of β found depends upon the identifiability constraints imposed. In Chapter 4, we argued that the choice of identifiability constraints is arbitrary, and no one set of identifiability constraints is preferable on statistical grounds to any other. We also know that the transformations in Equations 6.4, 6.5 and 6.6 allow us to change between different, equivalent sets of parameters (i.e., different arbitrary identifiability constraints) without changing the historical fit to data, whilst using well-identified projection processes for the period and cohort parameters means that the arbitrary choice of identifiability constraints will not affect the projected mortality rates. We therefore propose the following approach.

First, we fit the model as in Section 6.3, imposing the constraints in Equations 6.7, 6.8 and 6.9. These constraints are convenient when fitting the model as they are simple to apply (by regressing the cohort parameters on the relevant deterministic trends) and do not depend upon what time series process we subsequently use to project the period and cohort parameters.

Second, we select an appropriate time series process for the cohort parameters, working backwards from our desired demographic significance for the parameters and the need for the process to be well-identified, as discussed in Chapter 4. For illustrative purposes,

we select the AR(1) around quadratic drift process discussed above.[21]

Third, we fit an AR(1) around quadratic drift to the historical cohort parameters. In doing so, we find $\beta = \left(-5.05 \times 10^{-4}, \quad -1.24 \times 10^{-5}, \quad -2.49 \times 10^{-7}\right)$. Numerically, these regression coefficients are small, however it is important to note that they are not equal to zero. In the long run, therefore, the small quadratic trend in the cohort parameters will result in the projected cohort parameters diverging significantly from zero, which conflicts with our desired demographic significance.

However, the magnitude of β is entirely dependent upon the identifiability constraints used, i.e., even if β is small, we see from Equation 6.20 that $\hat{\beta} = \beta + A$ can be arbitrarily large depending upon the value of A. Therefore, any decision to ignore β would also be entirely dependent upon the arbitrary identifiability constraints. Thus, we are unable to test β and set it to zero if it proves statistically insignificant, since the results of any statistical tests on them would also depend upon the arbitrary identifiability constraint. Hence, the choice of time series to use for γ_y cannot be motivated by arguments based on statistical significance or goodness of fit, but must be determined by the identifiability issues present in the model, in order to avoid generating poorly-identified projections of mortality rates that depend on the arbitrary constraints imposed when fitting the model.

Since the value of β depends upon the identifiability constraints, we can work backwards to impose $\beta = 0$ by choosing a new set of identifiability constraints. To do this, we use the transformations in Equations 6.4, 6.5 and 6.6, with the values of the free parameters in these transformations given by the fitted values of β found above. This gives an equivalent set of historical parameters, with the original constraints in Equations 6.7, 6.8 and 6.9 over-ridden by the new constraint, $\beta = 0$. Imposing $\beta = 0$ in this fashion does not change our fitted mortality rates (as it merely involves using the invariant transformations), nor does it affect the projected mortality rates, since all the time series processes used for the period and cohort parameters are well-identified. However, it will ensure that our projected cohort parameters have the subjective demographic significance we desire for them from Chapter 2, namely that they lack deterministic trends.

The identifiability constraint $\beta = 0$ could not have been imposed when fitting the model to data, since it depends on knowing which time series process we would use to project

[21] However, in Section 6.5.3, we will extend this using a Bayesian approach to allow for the issues discussed in 6.5.1.

the cohort parameters a priori.[22] It therefore makes sense - and is certainly more convenient - to use the original set of identifiability constraints (Equations 6.7, 6.8 and 6.9), to fit the model to data and analyse the fitted cohort parameters. Once we have done this and chosen an appropriate time series process to project the cohort parameters, the fitting constraints can be revisited and we can switch to the more convenient set of identifiability constraints for projecting the model. Because all sets of fitted parameters give the same fitted mortality rates, and because using well-identified projection methods for both the period and cohort parameters means that, when we project any of these sets of parameters, we obtain the same projected mortality rates, we are free to switch between them at any stage of the analysis depending on which set of identifiability constraints is most convenient at the time. This is discussed in depth in Chapter 4.

6.5.3 A Bayesian approach for projecting the cohort parameters

From Section 6.5.1, we see that we must be careful when allowing for the uncertainty in the cohort parameters, as our estimates to date will be based only on incomplete information. In attempting to allow for this uncertainty, it therefore makes sense to develop a process that is consistent with the nature of our observation of each cohort.

We do this using a Bayesian technique, since Bayesian methods are well suited to allowing for the inherent uncertainty in parameter estimates based on partial information, but there are prior views regarding the process generating the data. Bayesian methods have been used extensively in order to fit various mortality models to data, for instance in Pedroza (2006), Cairns et al. (2006b), Reichmuth and Sarferaz (2008) and Mavros et al. (2014), often using Markov chain Monte Carlo (MCMC) techniques. However, they have not been used to model the underlying processes generating the cohort parameters. Accordingly, the fitted values of γ_y from models with cohort parameters fitted using MCMC techniques will suffer from exactly the same issues as those described in Section 6.5.1. Instead, we construct a Bayesian framework for the cohort parameters from the ground up, starting by specifying the underlying data generating process of each individual cohort parameter and then incorporating a (well-identified) time series

[22] In principle, if the final time series processes are known in advance or determined by a trial two step sequential estimation of the model and time series processes, it is possible to fit the model and time series processes to data jointly in a one step process. This can be done either using maximum likelihood techniques (as in Dowd et al. (2011b), or Bayesian Markov chain Monte Carlo techniques, as in Pedroza (2006). However, such techniques are complicated to implement and so are not practical when using sophisticated mortality models or if the model is intended to be used for different datasets, where different time series processes might be appropriate.

process governing the evolution of the cohort parameters across years of birth.

6.5.3.1 The data generating process

We start by noting that our dataset gives us a limited number of observations for each cohort, each of these observations giving us a small amount of information regarding the mortality effects specific to that cohort. We also note that the value of each observation is proportional to the fraction of the cohort which dies at that age, with ages with many deaths giving relatively more insight than ages experiencing few deaths. We formalise this intuition as follows.

Consider a cohort born in year y where a proportion, d_x, of the total cohort dies at age x (assuming ages in the range $[1, X]$ and no other decrements from the population other than death, such as migration). For simplicity, d_x is assumed to be the same for all cohorts.[23] Therefore, by the time the cohort has reached age x, we have seen a proportion $D_x = \sum_{\xi=1}^{x} d_\xi$ of the cohort die. Trivially, $D_X = \sum_{\xi=1}^{X} d_\xi = 1$.

We start by assuming that each observation of cohort y at age x gives us a packet of information, γ_y^x, relating to the cohort-specific mortality effects. We assume

$$\gamma_y^x | \Gamma_y, \sigma^2 \sim N\left(\Gamma_y, \frac{\sigma^2}{d_x}\right) \tag{6.21}$$

where Γ_y is the common mean of the information packets for year of birth y. We assume that the information packets are conditionally independent of each other, apart from sharing a common mean. This implies that an observation of a cohort at age 50 only depends upon the observation of the same cohort aged 40 via the mean, Γ_y, and so observations of the γ_y^x can be used to estimate this unknown variable. We will assume a prior distribution for Γ_y based on the time series structure for the cohort parameters considered in Section 6.5.3.2.

What we are primarily interest in, however, is the "ultimate" cohort parameter, γ_y. This is the lifelong mortality effect experienced by the cohort, and is constructed from the packets of information observed at each age. Because the ultimate cohort parameter is a lifelong effect, it will only be known fully at the extinction of the cohort (i.e., at time

[23] In practice, we take d_x to be given by the fitted mortality rates in the final year of the data. However, the results are relatively insensitive to the choice of d_x as long as these reflect a plausible pattern of deaths from a cohort across different ages.

$y + X$), but will be unobservable at any time before this. We assume that the ultimate cohort parameter is given by the weighted sum of the information packets, with the weights given by the schedule of deaths for the cohort, i.e.,

$$\gamma_y = \sum_{x=1}^{X} d_x \gamma_y^x \tag{6.22}$$

From this, we find the distribution of the ultimate cohort parameter, assuming we have observed no information packets to date (e.g., for cohorts which have yet to be born)

$$\gamma_y | \Gamma_y, \sigma^2 \sim N(\Gamma_y, \sigma^2) \tag{6.23}$$

Thus, Γ_y is also the mean of the ultimate cohort parameter, as well as the mean of the information packets. Note that the packets are all a lot more variable than the ultimate cohort parameter, since d_x will tend to be small (of the order of a few percent of people in a cohort dying at each age).

Before the extinction of the cohort, γ_y is unobservable and we will have only partial information regarding the cohort, based on the packets of information observed to date. The challenge, therefore, is to find the distribution of the ultimate cohort parameter given the partial information we have at time t. We will typically assume that t is fixed at the current year of observation (i.e., the last year of the dataset).[24] At this time, we have received the first $t - y$ packets of information, i.e., $\gamma_y^x,\ x \in [1, t-y]$. We, therefore, define the partial sum of the packets, $\underline{\gamma}_y(t) = \sum_{x=1}^{t-y} d_x \gamma_y^x$. The distribution of this partial sum is given by

$$\underline{\gamma}_y(t) | \Gamma_y, \sigma^2 \sim N\left(D_{t-y}\Gamma_y, D_{t-y}\sigma^2\right) \tag{6.24}$$

Unlike the individual information packets, γ_y^x, the partial sums, $\underline{\gamma}_y(t)$, are, in principle, observable at time t and could be found from the available data . However, they are not the same as the estimated cohort parameters found when fitting a mortality model to the available data at time t. This is because the expected value of the partial sums depends upon D_{t-y}, i.e., the proportion of the cohort expected to have died to date, and so we observe very small values of $\underline{\gamma}_y(t)$ for cohorts which have just been born, but considerably larger values for older cohorts (for fixed Γ_y). This is inconsistent with the assumption, implicit in the majority of APC mortality models, that the cohort parameters have the

[24] In Chapter 12, this is relaxed and the year of observation is allowed to change to reflect the impact of new observations on the previously estimated cohort parameters.

same scale.[25]

Therefore, we define "interim" cohort parameters, $\overline{\gamma}_y(t) = \frac{1}{D_{t-y}}\underline{\gamma}_y(t)$. From Equation 6.24, we see that the $\overline{\gamma}_y(t)$ have distribution

$$\overline{\gamma}_y(t)|\Gamma_y, \sigma^2 \sim N\left(\Gamma_y, \frac{1}{D_{t-y}}\sigma^2\right) \tag{6.25}$$

Not only do the $\overline{\gamma}_y(t)$ have means independent of D_{t-y}, but they have variances which are inversely proportional to the number of deaths expected from the cohort to date, which is consistent with Equation 6.17 and the analysis of Section 6.5.1. Therefore, we identify the interim cohort parameters, $\overline{\gamma}_y(t)$, with the cohort parameters estimated by the model in Section 6.3 and shown in Figure 6.4. Hence, we are able to obtain values of $\overline{\gamma}_y(t)$ by fitting the APC model to data. The interim cohort parameters, $\overline{\gamma}_y(t)$ are assumed to be known at time t, as opposed to having the distribution in Equation 6.25, and similarly the partial sums, $\underline{\gamma}_y(t)$, are also assumed to be known at time t. It is trivial to move between the fitted $\overline{\gamma}_y(t)$ and the partial sums, $\underline{\gamma}_y(t)$, which are more fundamental in the analysis.

We can use the knowledge of $\overline{\gamma}_y(t)$ (and $\underline{\gamma}_y(t)$) to update the distribution for the ultimate cohort parameter, γ_y by conditioning on the partial information we have to time t. To do this, we note that, for times in the interval $y \leq t < y + X$

$$\begin{aligned}\gamma_y &= \sum_{x=1}^{t-y} d_x \gamma_y^x + \sum_{x=t-y+1}^{X} d_x \gamma_y^x \\ &= \underline{\gamma}_y(t) + \sum_{x=t-y+1}^{X} d_x \gamma_y^x\end{aligned} \tag{6.26}$$

Therefore, from Equation 6.21, we find

$$\gamma_y|\underline{\gamma}_y(t), \Gamma_y, \sigma^2 \sim N(\underline{\gamma}_y(t) + (1 - D_{t-y})\Gamma_y, (1 - D_{t-y})\sigma^2) \tag{6.27}$$

Thus, we have found the distribution of the ultimate cohort parameters for year of birth y, conditional on our observations of the cohort to date and its prior expected value. However, we have not made any assumptions regarding the form that this prior expectation should take and, in particular, how this expected value relates to the values for other neighbouring cohorts.

[25] This is a consequence of having a simplified age/cohort structure and setting $\beta_x^{(0)} = 1$ discussed in Chapter 2.

6.5.3.2 Time series dynamics

The dependence of the ultimate cohort parameters, γ_y, upon the preceding cohorts is given by the time series process driving the dynamics of the cohort parameters. These assumed time series dynamics act as a prior distribution in the Bayesian approach. Working backwards from our desired demographic significance for the cohort parameters, we said in Section 6.5.2, that an AR(1) process around a quadratic drift can provide projections in line with our desire for stationary but persistent cohort parameters relatively parsimoniously. Writing the AR(1) process around a quadratic drift in distributional terms gives

$$\gamma_y|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}), \sigma^2\right) \tag{6.28}$$

Comparing this with Equation 6.23, we see that using the AR(1) process around a quadratic drift is equivalent to setting $\Gamma_y = \beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})$.[26] This choice for Γ_y also feeds through into the distributions both of the partial sums, $\underline{\gamma}_y(t)$, in Equation 6.24 to give

$$\underline{\gamma}_y(t)|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(D_{t-y}(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})), D_{t-y}\sigma^2\right) \tag{6.29}$$

and of the information packets, γ_y^x, in Equation 6.21 to give[27]

$$\gamma_y^x|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}), \frac{\sigma^2}{d_x}\right) \tag{6.30}$$

To incorporate both sources of information regarding the ultimate cohort parameter, γ_y (i.e., the partial information observed to date for the cohort and that from the cohort parameter for the previous year of birth using the time series structure), we substitute the expression for Γ_y into Equation 6.27, to obtain

$$\begin{aligned}\gamma_y|\underline{\gamma}_y(t),&\gamma_{y-1},\beta,\rho,\sigma^2 \sim \\ &N\left(\underline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})), (1 - D_{t-y})\sigma^2\right)\end{aligned} \tag{6.31}$$

This expression gives the distribution of the ultimate cohort parameter for cohort y, given our observations of the cohort parameter to date and the previous ultimate cohort parameter, γ_{y-1}. It can, therefore, be considered as the posterior distribution in the Bayesian approach, since it takes the prior distribution given by the time series dynamics

[26] The model could, theoretically, be extended to allow for more lags and an AR(p) structure via a different choice for Γ_y.

[27] While the distribution for γ_y^x is not used here, it is necessary when updating the estimates of the cohort parameters for additional data, as done in Chapter 12.

in Equation 6.28 and updates it by incorporating the information observable in $\underline{\gamma}_y(t)$. This posterior distribution can be used for simulation purposes, especially when it is rewritten in the form

$$\begin{aligned}\gamma_y &= \underline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1})) + \epsilon_y \\ \epsilon_y &\sim N(0, (1 - D_{t-y})\sigma^2)\end{aligned} \tag{6.32}$$

We refer to this as the "updating equation", which we can use to simulate sample paths for the ultimate cohort parameters, γ_y, over the range $t - X < y < Y$ (where Y is the last cohort in the data for which we have estimated a cohort parameter).

If we were to write Equation 6.32 using the interim cohort parameters, $\overline{\gamma}_y(t)$, estimated by the model, instead of the partial sums, $\underline{\gamma}_y(t)$, we can see that the expectation of the ultimate cohort parameter is of the form of a weighted sum of the fitted parameter based on observations of the cohort to time t and the expected value from the time series dynamics

$$\mathbb{E}\gamma_y|\underline{\gamma}_y(t), \gamma_{y-1}, \beta, \rho, \sigma^2 = D_{t-1}\overline{\gamma}_y(t) + (1 - D_{t-y})(\beta X_y + \rho(\gamma_{y-1} - \beta X_{y-1}))$$

In this form, the approach can be compared to a "credibility analysis" of the cohort parameters as discussed in Chapter 7 of Kaas et al. (2001), since our estimate of the true parameter is formed as a weighted average of our observed parameter and what would be predicted by the time series. These weights, i.e., the proportion of each cohort expected to have died by the observation date, are shown in Figure 6.5. We can see that we place a high degree of confidence in our estimates of the cohort parameters before c. 1930 (i.e., individuals currently aged around 80), but this falls rapidly for younger cohorts. For these, the second term in Equation 6.32 will dominate.

While useful for simulation purposes, Equation 6.31 is not the end of the story, since it is still conditional on knowing the previous ultimate cohort parameter, γ_{y-1}. However, for the majority of cohort parameters, the previous ultimate cohort parameter will also be unknown at time t. However, it is possible to solve Equation 6.31 iteratively to remove the dependence on γ_{y-1} and obtain the distribution for the cohort parameter γ_y at time t, based solely on the observations made to date. We do this by writing

$$\gamma_y|\mathcal{F}_{t,y}, \beta, \rho, \sigma^2 \sim N(M(y,t), V(y,t)) \tag{6.33}$$

where $\mathcal{F}_{t,y}$ represents the sum total of information known at time t about cohorts up to and including year of birth y, i.e., $\{\underline{\gamma}_\upsilon(t)\ \upsilon \leq y\}$, and $M(y,t)$ and $V(y,t)$ are the mean

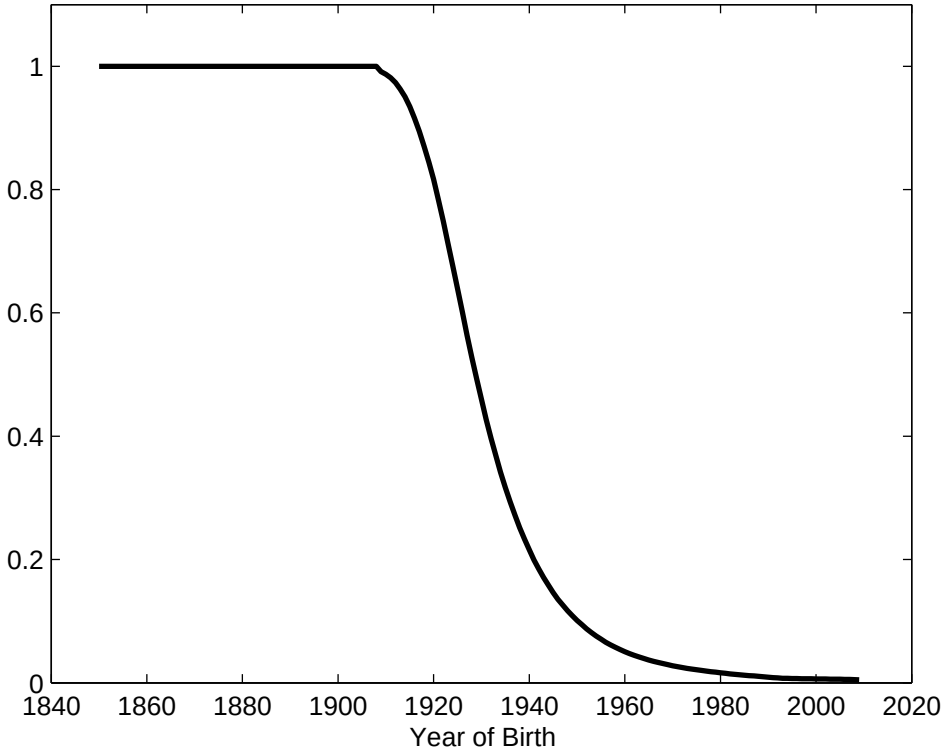


FIGURE 6.5: Deceased proportion of cohort, D_y

and variance functions, respectively. From Equation 6.31 and Bayes Theorem, we work backwards to give

$$\begin{aligned}\gamma_y|\mathcal{F}_{t,y},\beta,\rho,\sigma^2 \sim N\Big(&\underline{\gamma}_y(t)+(1-D_{t-y})(\beta X_y+\rho(M(y-1,t)-\beta X_{y-1})),\\ &(1-D_{t-y})\sigma^2+(1-D_{t-y})^2\rho^2V(y-1,t)\Big)\\ \Rightarrow M(y,t)=&\underline{\gamma}_y(t)+(1-D_{t-y})(\beta X_y+\rho(M(y-1,t)-\beta X_{y-1})) \qquad (6.34)\\ V(y,t)=&(1-D_{t-y})\sigma^2+(1-D_{t-y})^2\rho^2V(y-1,t) \qquad (6.35)\end{aligned}$$

This gives us iterative equations for the mean and variances functions, respectively, for the ultimate cohort parameters based on the information observed to date, which can be solved to give

$$M(y,t)=\sum_{s=0}^{\infty}\left[\prod_{r=0}^{s-1}(1-D_{t-y+r})\right]\rho^s\left[\underline{\gamma}_{y-s}(t)+(1-D_{t-y+s})\beta(X_{y-s}-\rho X_{y-s-1})\right] \qquad (6.36)$$

$$V(y,t)=\sum_{s=0}^{\infty}\left[\prod_{r=0}^{s-1}(1-D_{t-y+r})^2\right](1-D_{t-y+s})\rho^{2s}\sigma^2 \qquad (6.37)$$

in closed form. We adopt the convention that empty products equal unity (i.e., $\prod_{r=0}^{s-1}(1-$

$D_{t-y+r}) = 1$ for $s = 0$). It is also important to note that, although these are written as infinite sums, they will in fact terminate as $D_X = 1$.

So far, this analysis has assumed that we know the parameters of the underlying time series dynamics, i.e., Equation 6.33 is conditional on knowing the values of β, ρ and σ^2. In practice, these parameters can be estimated from the fitted cohort parameters, once we find the predictive distribution for $\underline{\gamma}_y(t)|\mathcal{F}_{t,y-1}$, i.e., the observed $\underline{\gamma}_y(t)$, given all previous $\underline{\gamma}_v(t)$. This can be calculated using Bayes Theorem and Equation 6.29 to give

$$\begin{aligned}\underline{\gamma}_y(t)|\mathcal{F}_{t,y-1}, \beta, \rho, \sigma^2 \sim N\,(&D_{t-y}(\beta X_y + \rho(M(y-1,t) - \beta X_{y-1})),\\ &D_{t-y}\sigma^2 + \rho^2 D_{t-y}^2 V(y-1,t)\big)\end{aligned} \tag{6.38}$$

This predictive distribution gives us the distribution of an observable quantity, $\underline{\gamma}_y(t)$, in terms other observable quantities, $\underline{\gamma}_v(t)$ (in $M(y,t)$), and the unknown time series parameters. This means that we can use quasi-maximum likelihood methods to estimate β, ρ and σ^2. As discussed in Section 6.5.2, in general, we will observe non-zero values for β, which is undesirable given our demographic significance for the cohort parameters. We, therefore, use the invariant transformations in Equations 6.4, 6.5 and 6.6 to set $\beta = 0$, as discussed in Section 6.5.2. This also has the benefit of simplifying both the expression for $M(y,t)$ in Equation 6.36 and the projections of the cohort parameters considerably.

So far, we have only considered the situation where we have two sources of information for each cohort, the observations to date and the time series structure. In order to project the cohort parameters into the future (i.e., beyond year of birth Y), we do not have any observations to date and therefore we simply use the AR(1) structure to generate projections. To project beyond the last fitted cohort parameter (assumed to be known for the time being), the AR(1) process gives

$$\gamma_{Y+\eta}|\gamma_Y, \rho, \sigma^2 \sim N\left(\rho^\eta \gamma_Y, \frac{1-\rho^{2\eta}}{1-\rho^2}\sigma^2\right)$$

To remove the dependence on γ_Y, which will be unknown in practice, we use Bayes Theorem to obtain

$$\gamma_{Y+\eta}|\mathcal{F}_{t,Y} \sim N\left(\rho^\eta M(Y,t), \frac{1-\rho^{2\eta}}{1-\rho^2}\sigma^2 + \rho^{2\eta} V(Y,t)\right) \tag{6.39}$$

The variance of this contains two parts. First, the variability from projecting the time series, which increases to a constant $\sigma^2(1-\rho^2)^{-1}$ as $\eta \to \infty$ as expected. Second, there

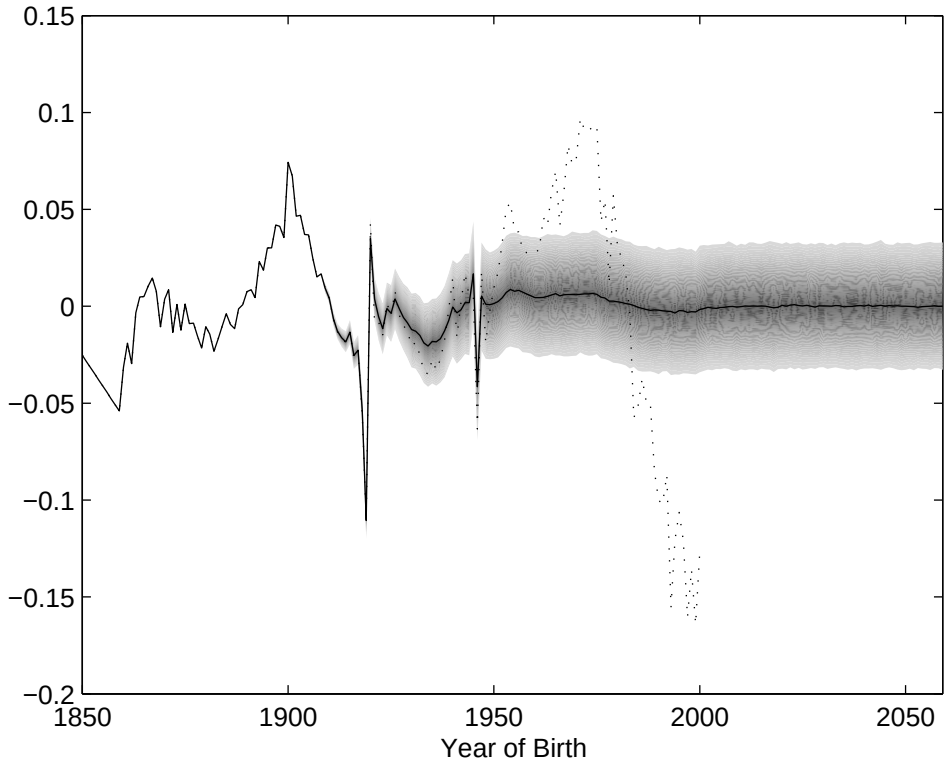


FIGURE 6.6: 95% fan chart of the projected cohort parameters using the Bayesian approach

is the variability from the fact that our initial value γ_Y is unknown: this source of variability decays exponentially. However, as $V(Y,t) < \sigma^2(1-\rho^2)^{-1}$,[28] this means that our confidence intervals for $\gamma_{Y+\eta}$ increase with time towards a limit.

As with Equation 6.31, it is helpful to rewrite Equation 6.39 in the form of an updating equation

$$\gamma_{Y+\eta} = \rho\gamma_{Y+\eta-1} + \varepsilon_y$$
$$\varepsilon_y \sim N(0, \sigma^2)$$

which can be used for generating sample paths. Again, we see that this is simply the time series process for an AR(1) process and is similar to Equation 6.32, but with $D_{t-y} = 0$ and $\beta = 0$, i.e., we are forecasting cohorts for which there have been no observed deaths to date.

[28] Mathematically, this is a consequence of $D_{t-Y} > 0$. More intuitively, it can be seen that $\sigma^2(1-\rho^2)^{-1}$ is the variability of a cohort parameter under the prior distribution from the AR(1) time series without any additional information from the data to refine the parameter estimate.

Figure 6.6 shows a fan chart of the values of the cohort parameters using this method, with the fitted parameters indicated by a dotted line for comparison. We note that the cohort parameters have three regimes:

1. $y \leq t - X$ (i.e., $y \leq 1909$): our data has a complete set of observations regarding the cohort and therefore we do not have any uncertainty in the cohort parameters (i.e., $\gamma_y = \underline{\gamma}_y(t) = \overline{\gamma}_y(t)$).

2. $t - X < y \leq Y$ (i.e., $y \in [1910, 1999]$):[29] we have partial observations for each cohort and, therefore, γ_y is not known with certainty but is constructed from the observations to date and the time series dynamics. However, older cohorts are considerably less variable as we have a greater number of observations for these years of birth (and observations including ages where a larger proportion of the cohort is expected to die). In contrast, the uncertainty in the parameter estimates grows rapidly for more recent cohorts.

3. $Y < y$ (i.e., $y \geq 2000$): we have no observations for these years of birth and so the projected cohort parameters are based solely upon the time series dynamics assumed.

It is important to note that, despite the qualitative differences between these three regimes, the confidence interval showing the uncertainty in the parameters blends smoothly between the fitted and the projected parameters, with no sharp discontinuity at the regime boundary. This is in contrast to the classical approaches discussed in Section 6.5.1, which would have the uncertainty of the cohort parameters increase sharply at the boundary between estimated cohort parameters, $y \leq Y$ (assumed known) and projected cohort parameters, $y > Y$ (projected using the time series). This is important in many applications, such as projecting annuity values, as discussed in Section 6.6.3, and also for valuing longevity-linked securities, as discussed in Chapter 8.

We also note from Figure 6.6 that the expectation of the ultimate cohort parameter, $M(y,t)$ (given by the centre of the confidence interval in Figure 6.6), can be significantly different from the cohort parameters estimated from data to time t, $\overline{\gamma}_y(t)$. Since these estimated cohort parameters were fitted (along with the other parameters in the model) on the basis of maximising the goodness of fit to data, using the Bayesian approach will worsen the fit to the historical data. However, the reduction in the goodness of fit is

[29] As discussed in Chapter 5, we do not fit cohort parameters for the last 10 years of birth in the data, due to the lack of observations. Instead, these are linearly interpolated to zero to prevent them interfering with the age/period terms.

relatively marginal,[30] as the difference between the two is only significant for the most recent cohorts, for whom we have relatively little data to fit the model. This worsening of the goodness of fit is also more than compensated by the more plausible projections and increased allowance for uncertainty in these parameter estimates. In addition, the use of the Bayesian approach for the cohort parameters may appear inconsistent with the use of the other fitted age and period functions in the model. However, these other parameters are estimated over a wide range of years of birth and so are not significantly affected by the changes to the most recent years of birth caused by using the Bayesian approach for the cohort parameters.[31]

Finally, we also see that the pattern of the fitted cohort parameters shown in Figure 6.4 after 1950 (i.e., a rapid increase and then decrease in cohort mortality relative to the baseline) is smoothed out, since it is not based on sufficient observations to be credible. Therefore, using the Bayesian approach will tend to avoid the issues found in Cairns et al. (2011a), where distinctive patterns in the most recent cohort parameters lead to projected mortality rates which are not biologically reasonable.

In summary, we propose a new Bayesian approach for projecting the cohort parameters, which involves updating a prior distribution for them based on assumed time series dynamics with the partial observations we have for each cohort from the available data. This is similar conceptually to a creditability analysis of the form familiar to actuaries. In addition, we have ensured that these projections are well-identified, in the sense that the projected mortality rates do not depend upon any arbitrary set of identifiability constraints imposed. Although this approach is complicated, it yields projections of the cohort parameters which we believe are more plausible and also allow for the uncertainty in the historical cohort parameters as we have only partial data regarding them.

6.6 Testing the projected mortality rates

Our aim is to develop techniques for projecting mortality rates that are more consistent with the features observed in the historical data and which make suitable allowance for longevity risk. This cannot be done by looking at the parameters of the model in isolation. Rather, we must look at the plausibility of the projected mortality rates and

[30] We find log-likelihoods of -3.09×10^{-4} using the estimated parameters and -3.25×10^{-4} using the expectation of the ultimate parameters, which is mainly due to worsening the fit to mortality data at age zero. This may indicate that the fitted cohort parameters attempt to overfit data at this unusual age, rather than capturing genuine lifelong mortality effects.

[31] In principle, the other age/period terms in the model could be re-estimated subsequent to determining $M(y,t)$. In practice, however, this was not done in this study.

associated indices in order to assess the reasonableness of the models developed. To test our projections, we follow the approaches of Dowd et al. (2010b) and Cairns et al. (2011a) by first backtesting the projection model to see if it could have predicted the mortality rates observed in the past, and then make longer term forecasts to assess the qualitative nature of the mortality forecasts.

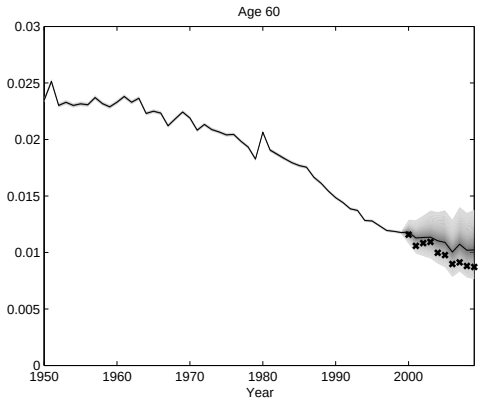
We combine the trend change model for the period functions and the Bayesian approach for the cohort parameters to project mortality rates into the future. We will call this the "consistent" approach since it has been designed to give projections which are consistent with the observed features of the historical data. For a comparison, we use an approach which simply uses a multivariate random walk for the period parameters and an AR(1) process for the cohort parameters. This approach is more typical of the projection methods used by previous studies, e.g. Cairns et al. (2006a), Cairns et al. (2009) and Haberman and Renshaw (2011). We will denote this the "naïve" approach, since, due to its simplicity, it is unable to give projections which are independent of the identifiability constraints, allow for structural breaks in the period functions, or allow for the uncertainty in the cohort parameters.

Our projections also allow for parameter uncertainty using the residual bootstrapping technique of Koissi et al. (2006). We also allow for idiosyncratic (Poisson) risk in the projected mortality rates when these are compared with the observed mortality rates in the backtesting exercise conducted in the following section.

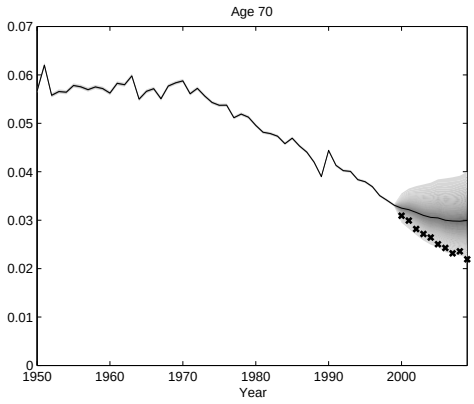
6.6.1 Backtesting the "consistent" and "naïve" approaches

We first test the consistent model using a backtesting procedure similar to that developed in Dowd et al. (2010b). The model is first fitted to data from 1950 to 1999 and then projected for the period 2000 to 2009. These projected mortality rates (allowing for both parameter uncertainty and idiosyncratic mortality risk) are then compared with the rates observed during this period. Results of this procedure at ages 60, 70 and 80 are presented in Figures 6.7 and 6.8 for the naïve and consistent approaches, respectively.[32] These show fan charts covering the 95% confidence interval for the projected mortality rates with crosses representing the observed mortality rates.

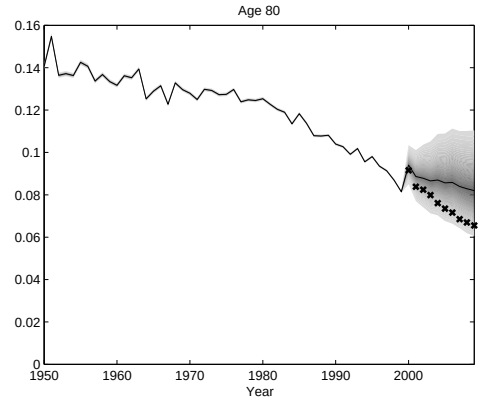
[32]These ages have been chosen as they are of greatest interest to annuity providers, such as life insurance companies and pension schemes, which are most affected by longevity risk. Similar figures for younger ages do not show any significant difference between the two models in terms of the ability to forecast mortality rates for the period 2000 to 2009.



(A) Age 60



(B) Age 70



(C) Age 80

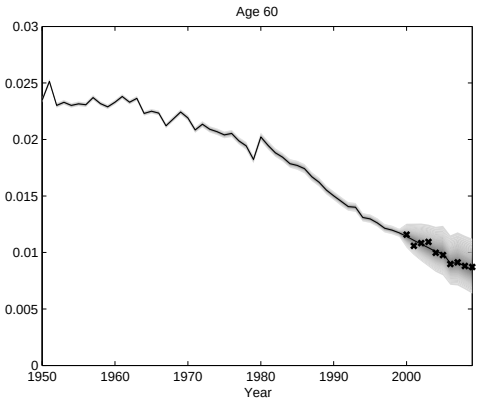
FIGURE 6.7: 95% confidence intervals for backtested mortality rates - Naïve approach

The fan charts show that at these key ages, the consistent approach gives considerably more accurate forecasts of mortality rates in comparison with the naïve approach. In particular, it is noted that the naïve projection method gives poor projections of mortality rates between ages 70 and 90 for more than five years ahead. Since it is these ages that are of most interest to providers of annuities and pension products and also where the numbers of deaths are greatest, this is of great concern.

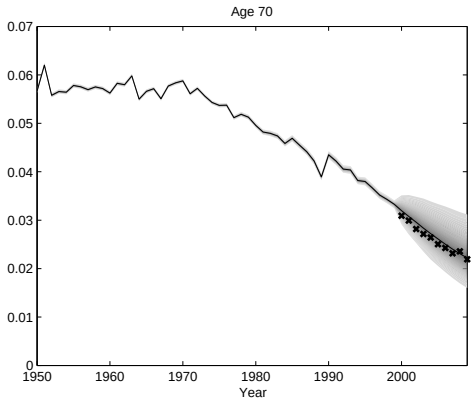
To test this statistically, we use the Dawid-Sebastiani scoring rule (DSS) discussed in Gneiting and Raftery (2007), as used in Riebler et al. (2012) and van Berkum et al. (2014).[33] To do this, we calculate the statistic

$$\text{DSS}_{x,t} = \frac{1}{5,000} \sum_{j=1}^{5,000} \frac{\left(\ln\left(\mu_{x,t}^{(j)}\right) - \ln\left(m_{x,t}\right)\right)^2}{\sigma_{x,t}} + \ln\left(\sigma_{x,t}^2\right) \tag{6.40}$$

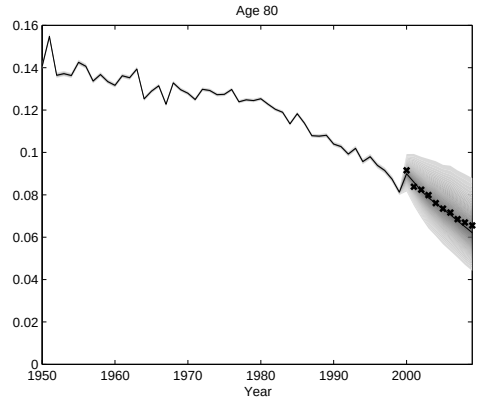
[33] We are indebted to Frank van Berkum for bringing this test to our attention.



(A) Age 60



(B) Age 70



(C) Age 80

FIGURE 6.8: 95% confidence intervals for backtested mortality rates - Consistent approach

where $\mu_{x,t}^{(j)}$ are the projected mortality rates for simulation j for age x and period t, $m_{x,t} = \frac{d_{x,t}}{E_{x,t}^c}$ are the observed mortality rates, and $\sigma_{x,t}$ is the standard deviation of the projected log mortality rates, estimated on the basis of 5,000 Monte Carlo simulations.[34] Thus, the Dawid-Sebastiani scoring rule gives a larger value if the observed mortality rates are a great distance from the centre of the confidence interval of the projected mortality rates, whilst taking into account the width of this confidence interval.

The difference between the DSS statistics using the consistent and naïve approaches at each age and period are shown in Figure 6.9. As can be seen, the consistent approach gives generally lower DSS statistics, indicating that the projected mortality rates are closer to those observed, for most ages and years, but especially at younger ages and

[34] We look at projected log mortality rates, unlike projected death counts as in van Berkum et al. (2014), since we expect these to be approximately normally distributed, and hence, Equation 6.40 is similar to a log-likelihood.

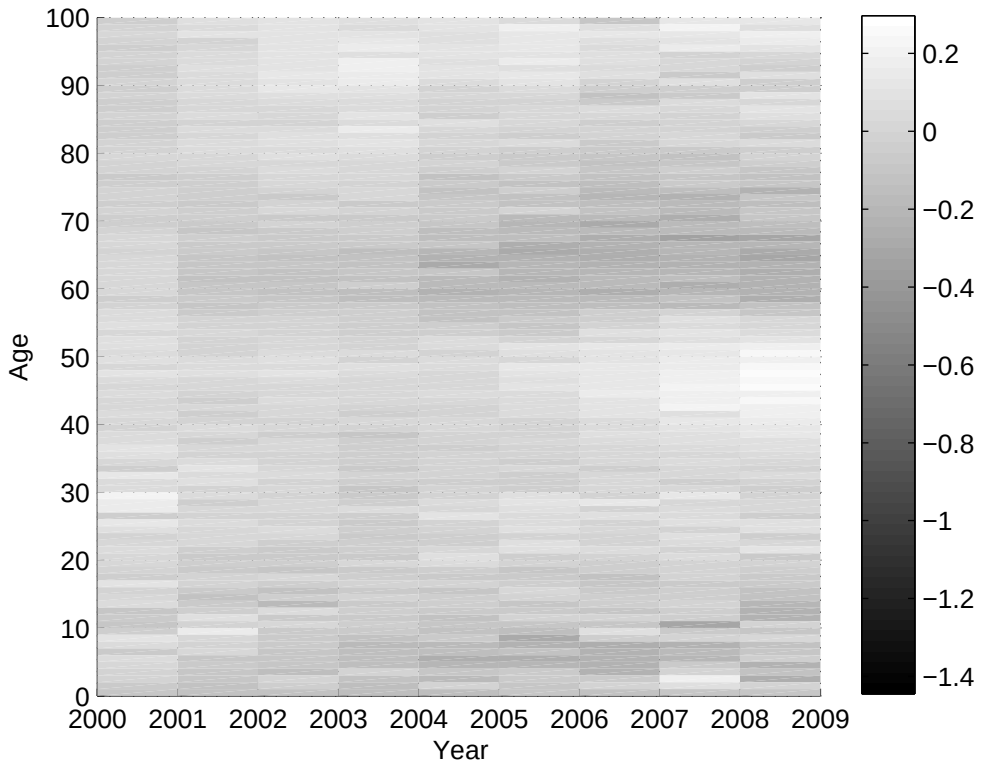


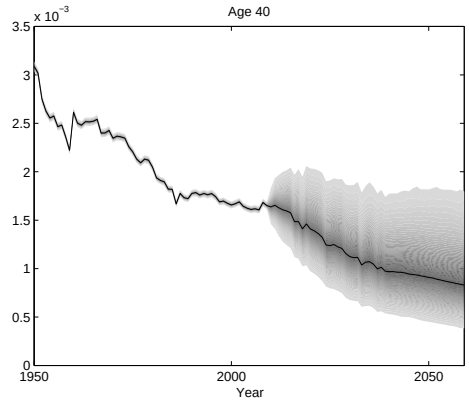
FIGURE 6.9: Heat map of differences in Dewid-Sebastiani score statistics between the consistent and naïve approaches across ages and projected years

ages 60 to 70. We can also calculated aggregate DSS statistics over all ages and years as

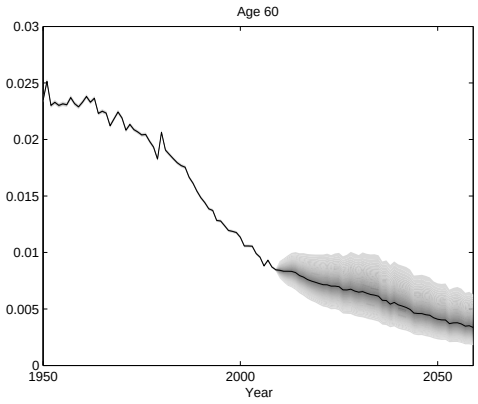
$$DSS = \frac{1}{101 \times 10} \sum_{x=0}^{100} \sum_{t=2001}^{2009} DSS_{x,t}$$

Doing this, we find aggregate DSS statistics of -3.80 for the consistent approach and -3.77 for the naïve approach, indicating that the consistent approach gives projections which are marginally closer to the observed mortality rates than the naïve approach overall. However, this statistic does not give greater weight to those ages of greatest interest (i.e., those at higher ages) and so should be used with caution.

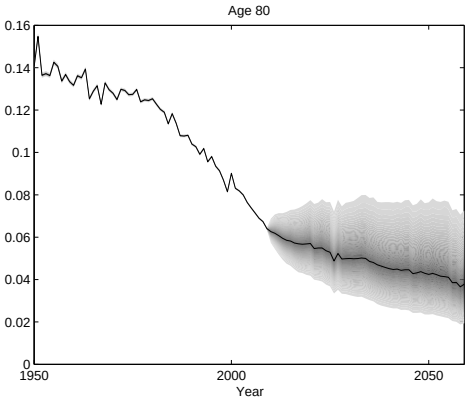
In summary, visual inspection of the backtesting exercise gives some evidence to suggest that the consistent approach gives more accurate projections of mortality rates, especially at the ages of greatest interest to pension and annuity providers. This is supported by the use of the Dawid-Sebastiani scoring rule to evaluate the closeness between the projected and observed mortality rates.



(A) Age 40



(B) Age 60



(C) Age 80

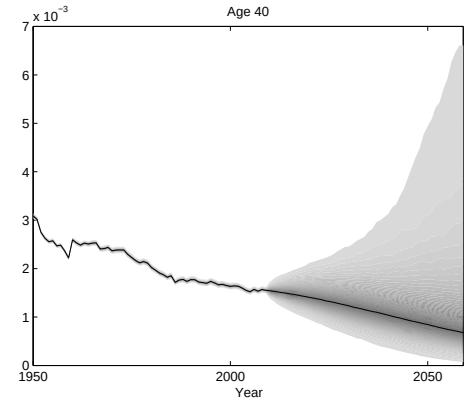
FIGURE 6.10: 95% fan charts of projected mortality rates - Naïve approach

6.6.2 "Consistent" and "naïve" mortality density forecasts

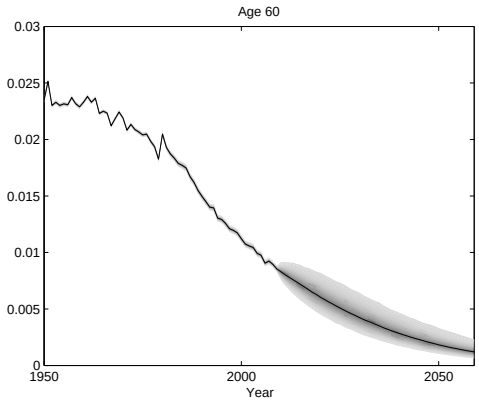
We use the consistent and naïve approaches to project mortality rates 50 years into the future. Figures 6.10 and 6.11 show projections for mortality rates at ages 40, 60 and 80 under these two alternative approaches.[35]

The first thing we note is that the naïve approach gives median projected mortality rates which are far less smooth than those given by the consistent approach. This is because they fully take account of the lack of smoothness in the fitted cohort parameters. In contrast, the Bayesian technique used in the consistent approach smooths the most recent cohort parameters via the prior time series model for them, leading to smoother median projected mortality rates overall, which might be felt to be more biologically reasonable.

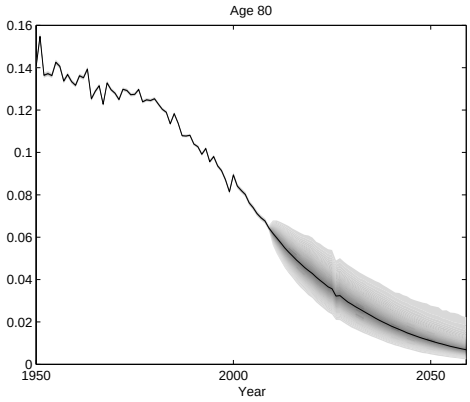
[35] These ages have been chosen as representative of the pattern of improvements across a broader range of ages than those shown in Figures 6.7 and 6.8.



(A) Age 40



(B) Age 60

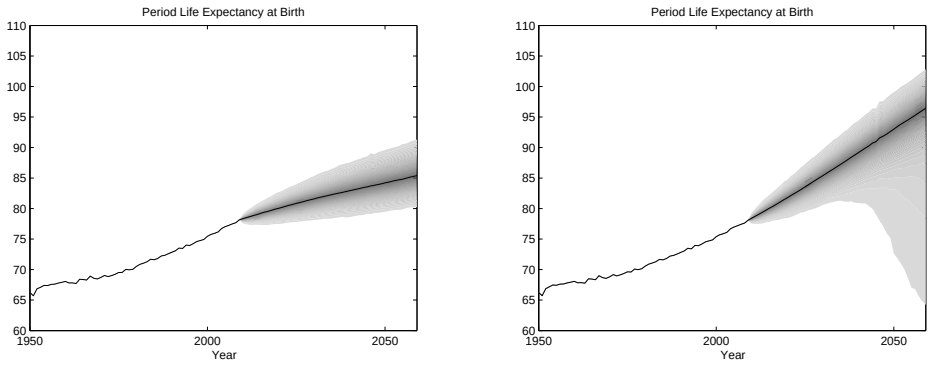


(C) Age 80

FIGURE 6.11: 95% fan charts of projected mortality rates - Consistent approach

Second, the consistent projection approach gives projected mortality rates which are considerably more variable than those using the naïve approach, especially at younger ages. For instance, Figure 6.10a shows that some tail scenarios have mortality rates for 40 year olds in 2060 in excess of those observed in the historical data, and comparable to those seen during the Second World War. This is mainly due to the potential for trend changes in $\kappa_t^{(3)}$. Allowing for these tail scenarios may appear extreme, but is desirable for consistency with the historical data, where these mortality rates have been more variable than those at higher ages. It is also consistent with our desire for biologically reasonableness as younger ages, which are typically subject to a wider range of significant causes of death, such accidents, suicides and disease pandemics (such as HIV or pandemic influenza) than older individuals. This means that our projection of mortality rates for these ages should be considerably more uncertain.

The differences between the two projection approaches also show up in the projections of aggregate measures of mortality, such as period life expectancy at birth as seen in Figure 6.12. In both cases, we see that our projection methods allow the high rates of



(A) Naïve approach (B) Consistent approach

FIGURE 6.12: 95% fan charts of projected period life expectancy at birth

increase in life expectancy observed in the recent past to continue in future. In the naïve approach, the rate of improvement in period life expectancy tails off in future. This is because the naïve approach uses a random walk with drift for $\kappa_t^{(1)}$, and so this parameter decreases roughly linearly. This leads to increases in life expectancies which get progressively slower and tail off as the entropy of the life table (as defined in Keyfitz (1985)) increases. In contrast, the consistent model uses a random walk with linear drift for $\kappa_t^{(1)}$ for reasons of identifiability, as discussed in Section 6.4.1. This gives projections for $\kappa_t^{(1)}$ which decrease faster than linearly, and these, in turn, have the consequence that life expectancies do not tail off in future but continue to increase at roughly the same rate as is currently observed. Projections using the consistent model therefore do not contradict the findings of Oeppen and Vaupel (2002) which show life expectancy increasing linearly over long time periods and predict that this linear increase in life expectancy will continue in future. Increases in life expectancy which do not tail off might also be considered to give more prudent (i.e., less financially optimistic) estimates of the long-term improvements in mortality rates for risk management purposes for annuity books, compared with models which implicitly assume that improvements in life expectancy tail off in future.

In addition, the higher variability of mortality rates at younger ages has the impact that the projected period life expectancy at birth under the consistent approach has asymmetric projection intervals. This is because improvements in mortality at younger ages have limited scope to reduce the number of deaths (which is already low) and so lengthen average life span, but deteriorations in mortality rates at younger ages have considerable scope to increase the number of premature deaths and so reduce life expectancy. However, in aggregate, we note that the extreme scenarios in Figure 6.12 show period life expectancy returning to a level last seen in 1950, which is not biologically unreasonable.

6.6.3 Risk management

One of the motivations in developing these techniques is to allow fully for the longevity risk present in benefits such as annuities. In respect of these benefits, a stylised life insurer will:

- hold reserves sufficient to meet a "best estimate" of the present value of the liabilities, and
- hold capital in excess of the reserves sufficient to cope with unexpected events up to a certain percentile in the probability distribution.

Figure 6.13 shows the "best estimates" of the present values of annuities for individuals aged 65 to 90 as the median projected annuity value using a real discount rate of 1% p.a. for both the consistent and naïve approaches. The consistent approach significantly increases the "reserves" required to back the annuities compared to the naïve approach, by between 5 and 15% depending on age. Mostly, this is due to $\kappa_t^{(1)}$ decreasing faster than linearly in future.

We can also look at the 95^{th} percentile of the distribution of present values to illustrate the additional "capital" required on top of these "reserves" using the consistent and naïve approaches.[36] In addition to increasing the reserves, the consistent approach increase the riskiness of the annuities slightly, requiring "capital ratios"[37] of about 11 to 13% depending on age, which is about 1 to 2% higher at ages below 80 than required using the naïve approach.

In summary, the consistent approach gives projected mortality rates which are more uncertain than the naïve approach, especially in the long term. This is of considerable importance for the providers of annuities, such as life insurance companies and pension schemes, as not allowing fully for the risk in these long-term projections may cause them to understate their capital requirements and reserves.

[36] For the avoidance of doubt, this is not directly comparable to the approach adopted in the calculation of the capital required under many modern solvency regimes, such as Solvency II (see EIOPA (2014)). This requires calculating the capital needed to protect against a change in the value of the reserves over a specified time period (usually one year) at a higher confidence level (e.g., the 99.5% level). Studies which look at this issue include Stevens et al. (2010), Plat (2011), Bauer et al. (2012) and Chapter 12.

[37] That is, $\frac{P_{95\%}-P_{50\%}}{P_{50\%}}$ where P_α is the α percentile of the distribution of annuity values.

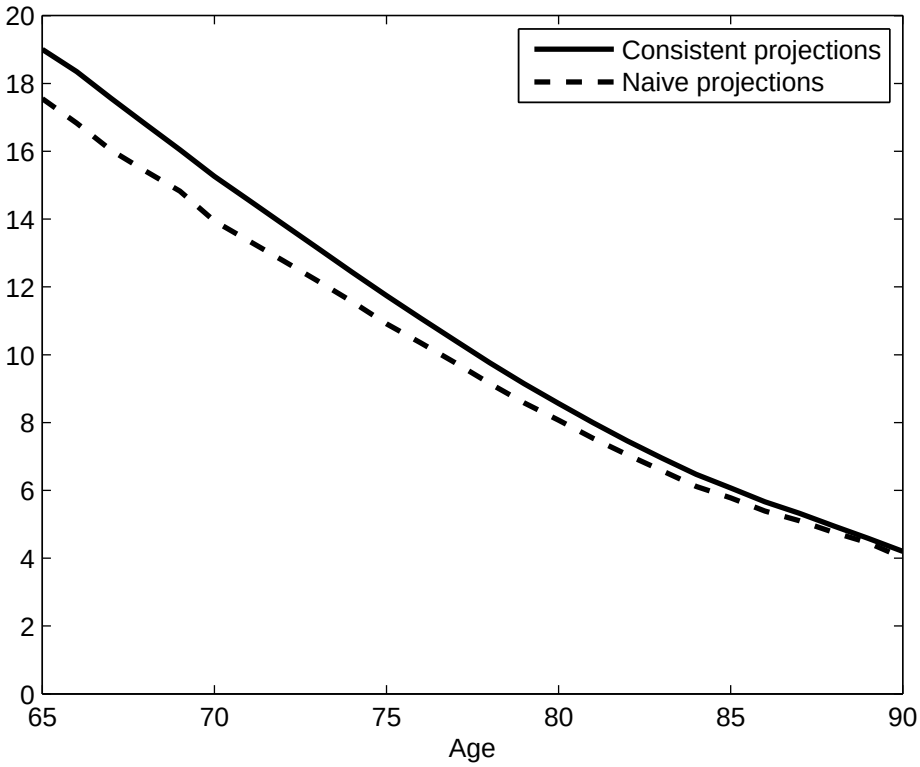


FIGURE 6.13: Expected present values of annuity using the naïve and consistent approaches (valued using 1% net discount rate)

6.7 Conclusions

The extrapolative approach to projecting mortality makes the core assumption that there is consistency between the evolution of mortality rates in the past and the future. We believe that assuming such consistency is practical and necessary. When using mortality models, there is a fundamental symmetry between the processes of fitting the model to historical observations to find parameter estimates and projecting parameter values to derive future observations. Therefore, we desire our projections of future patterns of mortality to be as similar to those observed in the historical data as possible.

In Chapter 5, we developed a "general procedure" for constructing mortality models, the purpose of which is to detect all of the statistically and demographically significant structure present in the data. In this study, we have developed techniques to make consistent projections into the future of many of the features found in the past, such as trend changes in the period functions. Further, we have allowed for uncertainty in our estimates of the cohort parameters due to the partial information we have observed to date for recent birth cohorts. Whilst these methods have been described in the context

of the model in Chapter 5, they can be easily adapted for any other age/period/cohort mortality model.

We also wanted to ensure that neither the model's fit to historical data nor its projections of future mortality rates are affected by the arbitrary choices made to identify the parameters in the model. To accomplish this, we had to utilise the results of Chapters 3 and 4 to select well-identified projection methods.

When using these techniques, we are able to obtain projections which are biologically reasonable and consistent with the most recent trends observed in the historical data. The techniques also suggest that standard approaches to projecting mortality rates may understate the risk inherent in projecting them into the future, as they do not fully allow for the possibility that features which have occurred in the past will recur in future. The past few decades have witnessed dramatic changes in mortality rates, and there is no evidence to suggest that the forthcoming decades will be different in terms of the magnitude of the pace of change or the challenges in making predictions. It is, therefore, vital that we make best use of our understanding of the past to incorporate sufficient uncertainty into our projections of what lies ahead.

6.A Forecast projection interval widths

Consider the model in Equation 6.16 with a constant drift subject to random changes in trend

$$\Delta\kappa_t = \mu_0 + \sum_{j=1}^{N} \nu_j I_{t\geq\tau_j} + \epsilon_t$$

We can solve this difference equation to give

$$\begin{aligned}\kappa_t &= \kappa_0 + \mu_0 t + \sum_{j=1}^{N} \nu_j (t-\tau_j)^+ + \sum_{s=1}^{t} \epsilon_s \\ &= \kappa_0 + \mu_0 t + \sum_{s=1}^{t} |\nu_s| J_s (t-s)^+ + \sum_{s=1}^{t} \epsilon_s\end{aligned}$$

where J_s is an indicator variable denoting whether a trend change occurred at time s and so takes the value $+1$ with probability $0.5p$ (corresponding to a positive trend change), -1 with probability $0.5p$ (corresponding to a negative trend change) and 0 with probability $(1-p)$ (no trend change). Assuming that the magnitudes of the trend changes $|\nu_s|$ are

independent and identically distributed, and are independent both of the process J_s (i.e., the direction of a trend change) and the innovations process ϵ_s, we have

$$\begin{aligned}
\mathbb{V}ar(\kappa_t) &= \mathbb{V}ar\left(\sum_{s=1}^{t}|\nu_s|J_s(t-s)\right) + \mathbb{V}ar\left(\sum_{s=1}^{t}\epsilon_s\right) \\
&= \sum_{s=1}^{t}(t-s)^2\mathbb{V}ar(|\nu_s|J_s) + \sigma^2 t \\
&= \sum_{s=1}^{t}(t-s)^2[(\mathbb{E}|\nu_s|)^2\mathbb{V}ar(J_s) + (\mathbb{E}J_s)^2\mathbb{V}ar(|\nu_s|) + \mathbb{V}ar(|\nu_s|)\mathbb{V}ar(J_s)] + \sigma^2 t \\
&= \sum_{s=1}^{t}(t-s)^2[p(\mathbb{E}|\nu_s|)^2 + p\mathbb{V}ar(|\nu_s|)] + \sigma^2 t \\
&= p\mathbb{E}|\nu|^2\sum_{s=1}^{t}(t-s)^2 + \sigma^2 t \\
&= p\mathbb{E}|\nu|^2(\frac{1}{6}t(t-1)(2t-1)) + \sigma^2 t
\end{aligned}$$

Therefore, for large t, $\mathbb{S}t\mathbb{D}ev(\kappa_t) \sim t^{1.5}$. This result is independent of the distribution of the trend change magnitudes $|\nu|$.

A similar result holds for the random walk with linear drift subject to random changes in trend. In this case, we find that

$$\kappa_t = \kappa_0 + (\mu_0 + 0.5\mu_1)t + 0.5\mu_1 t^2 + \sum_{s=1}^{t}|\nu_s|J_s(0.5((t-s)^2 + (t-s))) + \sum_{s=1}^{t}\epsilon_s$$

which can be solved in a similar fashion to give $\mathbb{S}t\mathbb{D}ev(\kappa_t) \sim t^{2.5}$ for large t.

Chapter 7

Identifiability, Cointegration and the Gravity Model

7.1 Introduction

Chapters 3 and 4 discussed the issue of identifiability in single population age/period/-cohort (APC) mortality models, and in particular how to obtain projections of mortality rates which do not depend upon the arbitrary identifiability constraints imposed.

Issues with identifiability in projections also exist if we project mortality for multiple populations rather than just one. Such multi-population projections are vital in order to allow for the correlations and dependencies between related populations that are influenced by similar biological and socio-economic drivers of changing mortality. It is essential that, in such a model, our projections do not depend on the arbitrary identifiability constraints imposed when fitting the model, but only on the underlying drivers of mortality evolution.

Many multi-population mortality models go beyond merely allowing for covariation between the stochastic evolution of mortality in different populations, and instead impose the stronger assumption of "coherence", i.e., that mortality rates in different populations should not diverge with time. Such an imposition is popular and intuitively appealing; however, we find that it usually cannot be imposed on a model in a fashion which does not depend on the arbitrary identifiability constraints. In addition, it can often lead to overriding the evidence from the historical data in order to impose our preconceptions

on projected mortality rates in a manner which we consider to be unscientific.

One model designed to achieve coherent projections of mortality between two populations is the "gravity model" of Dowd et al. (2011b). This model adopted a cointegration framework to project the period and cohort terms from the classic APC model fitted to each population. However, as originally formulated, the model is not well-identified, since the projections from it depend on the identifiability constraints imposed when fitting the classic APC model. Later work by Zhou et al. (2014) applied the framework of the gravity model to the period terms from the Lee-Carter model (Lee and Carter (1992)) and avoided some of the issues present in the original model of Dowd et al. (2011b). However, this new form of the model is still not well-identified, since it gives projections dependent upon the identifiability constraints imposed by the user.

In this study, we discuss the issue of identifiability in cointegration models and apply this to the specific context of the gravity model in order to obtain a well-identified model. Section 7.2 discusses the classic APC model which was used in Dowd et al. (2011b) to fit mortality rates in both populations. Section 7.3 outlines the gravity model introduced in Dowd et al. (2011b) and places it in the context of more general cointegration models. Section 7.4 discusses why the gravity model is not well-identified and how it can be modified to give well-identified projections. Section 7.5 discusses the model of Zhou et al. (2014), how it differs from the gravity model of Dowd et al. (2011b) and the issues with identifiability which are still present. Finally, Section 7.6 generalises these results to a broader class of mortality models and Section 7.7 concludes.

7.2 Identifiability in the classic APC model

The simplest APC model (referred to here as the "classic APC model") has a long history and is widely used in the fields of medicine, epidemiology and sociology as well as in demography and actuarial science. It has the form in Equation 7.1[1]

$$\ln(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \tag{7.1}$$

The parameters in the classic APC model cannot be estimated uniquely by reference to the data alone. A model is fully identified when all the parameters in it can be uniquely determined by reference to the available data. In contrast, the classic APC model is not

[1] In this chapter, we assume that ages, x, are in the range $[1, X]$ and periods, t, are in the range $[1, T]$ and therefore that years of birth, y, are in the range $(1 - X)$ to $(T - 1)$.

fully identified because there exist different sets of parameters which will give the same fitted mortality rates and consequently the same goodness of fit for any data set.

We can see that this model is not fully identified, since if we use the transformations in Equations 7.2, 7.3 and 7.4 to obtain new sets of parameters, we do not change our fit to the data (we call such transformations "invariant" for this reason)

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - a, \kappa_t + a, \gamma_y\} \tag{7.2}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x - b, \kappa_t, \gamma_y + b\} \tag{7.3}$$

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_y\} = \{\alpha_x + cx, \kappa_t - ct, \gamma_y + cy\} \tag{7.4}$$

Because different sets of parameters give the same fit to the data, we cannot use the data to choose between them. Typically, we impose identifiability constraints on the parameters in order to specify them uniquely. For instance, a commonly used set of identifiability constraints is $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and $\sum_y n_y \gamma_y (y - \bar{y}) = 0$.[2] We refer to these identifiability constraints as "natural", since they allow us to impose our interpretation of the demographic significance[3] of the parameters onto the model. For example, the first two of these constraints mean that α_x can be interpreted as an "average" level of mortality at age x over the period, with κ_t and γ_y representing deviations from this average level. The third constraint requires that there are no deterministic linear trends within the fitted cohort parameters, since any linear trend has been arbitrarily assigned to the age and period effects. This is in line with the demographic significance we assign to the cohort parameters in Chapter 2, namely that the cohort parameters should be centred around zero and should not show any long term trends. This means that cohort effects are interpreted as deviations in the mortality experienced by one cohort relative to that of adjacent years of birth.

However, it is important to note that these additional identifiability constraints, although having a natural interpretation, are arbitrary and ad hoc. While they might allow us to interpret the parameters in terms of their demographic significance, this interpretation nevertheless depends entirely on the user's judgement rather than on the underlying data. Of specific importance in the context of this study, Dowd et al. (2011b) used the

[2] Here n_y is the number of observations of cohort y in the data and so $\sum_y n_y \gamma_y = \sum_{x,t} \gamma_{t-x}$, and a bar denotes the arithmetic mean of the variable over the relevant data range, e.g., $\bar{y} = \frac{1}{X+T-1} \sum_y y = 0.5(X + T)$.

[3] Demographic significance is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

constraints $\sum_t \kappa_t = 0$, $\sum_y n_y \gamma_y = 0$ and a third constraint described in terms of minimising a tilting parameter δ, which can be written as $\sum_{x,t}(x - \bar{x})\gamma_{t-x} = 0$. The impact of using either the "natural" or the Dowd et al. (2011b) identifiability constraints when making projections is assessed in Section 7.4.3.

Since the identifiability constraint we choose to impose are arbitrary and do not affect the historical fitted mortality rates, they should also not affect the future projected mortality rates either. In consequence, we should obtain the same projected mortality rates for any set of identifiability constraints, including but not limited to the two discussed above. We say that models with this property are "well-identified".

7.3 The gravity model

The "gravity" model was introduced in Dowd et al. (2011b) in order to obtain mortality projections for two different populations which do not diverge with time.[4] This model might be appropriate for a small population, such as the lives in an annuity book or pension scheme, which is a subpopulation of a much larger population, such as a national population. The analogy the authors use is of the smaller population being like a planet in orbit around a star (the larger population).

The gravity model requires that the classic APC model of Equation 7.1 is fitted to two populations[5] and the period functions projected using

$$\begin{aligned} \kappa_t^{(I)} &= \nu^{(I)} + \kappa_{t-1}^{(I)} + \epsilon_t^{(I)} \\ \kappa_t^{(II)} &= \nu^{(II)} + \kappa_{t-1}^{(II)} + \phi(\kappa_{t-1}^{(I)} - \kappa_{t-1}^{(II)}) + \epsilon_t^{(II)} \end{aligned} \tag{7.5}$$

The parameter $\phi \in [0, 1)$ is designed to ensure that the difference, $\kappa_t^{(I)} - \kappa_t^{(II)}$, is stationary and, therefore, the period functions in the different populations do not diverge.

[4] This model is functionally equivalent to the model in Cairns et al. (2011b), which differs only in the presentation of the model and the techniques used to fit it to data. Therefore, the comments made in this chapter for the gravity model are also applicable to the model of Cairns et al. (2011b).

[5] In Dowd et al. (2011b), these were referred to as populations 1 and 2, with the period and cohort functions numbered accordingly. To avoid confusion with the different period functions $\kappa_t^{(i)}$ for models with more than one age/period term fitted to a single population, we shall refer to the populations as I and II and label the period functions $\kappa_t^{(I)}$ and $\kappa_t^{(II)}$ respectively.

We can rewrite Equation 7.5 as

$$\Delta \begin{pmatrix} \kappa_t^{(I)} \\ \kappa_t^{(II)} \end{pmatrix} = \begin{pmatrix} \nu^{(I)} \\ \nu^{(II)} \end{pmatrix} + \begin{pmatrix} 0 \\ \phi \end{pmatrix} \begin{pmatrix} 1, & -1 \end{pmatrix} \begin{pmatrix} \kappa_{t-1}^{(I)} \\ \kappa_{t-1}^{(II)} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{(I)} \\ \epsilon_t^{(II)} \end{pmatrix} \tag{7.6}$$

This model is just a special case of a more general cointegration model, although this interpretation was not commented upon in Dowd et al. (2011b). A number of papers have suggested or implemented cointegration as a means of projecting the period parameters of mortality models for different populations. Cointegration was first suggested in the work of Carter and Lee (1992), but was more recently used in the modelling of Li and Hardy (2011) and Yang and Wang (2013).

Cointegration between the period functions requires that we model the vector of time series processes as

$$\Delta \boldsymbol{\kappa}_t = \nu X_t + \sum_{i=1}^{p-1} \Gamma_i \Delta \boldsymbol{\kappa}_{t-i} + \Pi \boldsymbol{\kappa}_{t-p} + \boldsymbol{\epsilon}_t \tag{7.7}$$

The rank of the matrix Π is then tested in order to identify the number of cointegrating relationships between the period functions in the model. If it is of rank $r < N$ (the number of period functions in $\boldsymbol{\kappa}_t$), then Π can be decomposed as $\Pi = \alpha\beta^\top$, where α and β are $N \times r$ matrices to give the interpretation that the rows of $\beta^\top \boldsymbol{\kappa}_{t-p}$ represent r stationary cointegrating relationships between the different period functions. In order to use cointegration robustly, we need to ensure that any statements we make about the rank of Π are independent of our choice of identifiability constraints.

We can therefore see that the gravity model in Equation 7.6 has the same form as Equation 7.7, with $p = 1$, $X_t = \begin{pmatrix} 1 \end{pmatrix}$, $r = 1$, $\alpha = \begin{pmatrix} 0, & \phi \end{pmatrix}^\top$ and $\beta = \begin{pmatrix} 1, & -1 \end{pmatrix}^\top$. The prescribed form for β imposes that there is a stationary cointegrating relationship of the form $\kappa_t^{(I)} - \kappa_t^{(II)} = Z_t$, and so ensures that relative mortality rates will not diverge between the two populations, whilst the prescribed form for α allows the interpretation that population I is dominant and so has no dependence on population II.

A related process was used in Dowd et al. (2011b) to project the cohort parameters from the model. This can be written as

$$\Delta\begin{pmatrix}\gamma_y^{(I)}\\ \gamma_y^{(II)}\end{pmatrix} = \begin{pmatrix}\mu^{(I)}\\ \mu^{(II)}\end{pmatrix} + \begin{pmatrix}\alpha^{(I)} & 0\\ 0 & \alpha^{(II)}\end{pmatrix}\Delta\begin{pmatrix}\gamma_{y-1}^{(I)}\\ \gamma_{y-1}^{(II)}\end{pmatrix} + \begin{pmatrix}0\\ \phi\end{pmatrix}\begin{pmatrix}1, & -1\end{pmatrix}\begin{pmatrix}\gamma_{y-1}^{(I)}\\ \gamma_{y-1}^{(II)}\end{pmatrix} + \begin{pmatrix}\varepsilon_y^{(I)}\\ \varepsilon_y^{(II)}\end{pmatrix} \tag{7.8}$$

We can therefore see that this is also similar to the cointegration relationship in Equation 7.7.[6]

7.4 Identifiability in the gravity model

7.4.1 Period functions

The values of $\kappa_t^{(I)}$ and $\kappa_t^{(II)}$ are not uniquely identifiable by the classic APC model, but instead depend upon our choice of identifiability constraints. Equations 7.2 and 7.4 give us the freedom to add linear trends in time to either or both time series independently, i.e.

$$\begin{pmatrix}\hat{\kappa}_t^{(II)}\\ \hat{\kappa}_t^{(II)}\end{pmatrix} = \begin{pmatrix}\kappa_t^{(I)}\\ \kappa_t^{(II)}\end{pmatrix} + \begin{pmatrix}a^{(I)}\\ a^{(II)}\end{pmatrix} + \begin{pmatrix}c^{(I)}\\ c^{(II)}\end{pmatrix} t$$
$$\hat{\boldsymbol{\kappa}}_t = \boldsymbol{\kappa}_t + \boldsymbol{a} + \boldsymbol{c}t \tag{7.9}$$

However, this transformation, despite leaving the fitted mortality rates unchanged if we make the appropriate offsets to the static age functions and cohort parameters, fundamentally alters the cointegration relationship in Equation 7.6 since

$$\begin{aligned}\Delta\hat{\boldsymbol{\kappa}}_t &= \Delta\boldsymbol{\kappa}_t + \boldsymbol{c}\\ &= \boldsymbol{\nu} + \alpha\beta^\top\boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t + \boldsymbol{c}\\ &= \boldsymbol{\nu} + \boldsymbol{c} - \alpha\beta^\top(\boldsymbol{a} + \boldsymbol{c}t) + \alpha\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t\\ &= \hat{\boldsymbol{\nu}} - \alpha\beta^\top\boldsymbol{c}t + \alpha\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t\end{aligned}$$

The transformed time series has a deterministic linear term, $\alpha\beta^\top\boldsymbol{c}t$, which was not present in the original parameterisation. This means that the time series structure in

[6] There is a slight difference between Equation 7.8 and the standard form of the cointegration relationship in Equation 7.7, in that Equation 7.8 involves a stationary term in $\boldsymbol{\gamma}_{y-1} = \left(\gamma_{y-1}^{(I)}, \quad \gamma_{y-1}^{(II)}\right)^\top$ rather than $\boldsymbol{\gamma}_{y-2}$. This could be solved by rearranging Equation 7.8 using $A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\beta^\top\boldsymbol{\gamma}_{y-1} = (A + \alpha\beta^\top)\Delta\boldsymbol{\gamma}_{y-1} - \alpha\beta^\top\boldsymbol{\gamma}_{y-2}$ and redefining the matrix A. However, this solution involves losing the particular structure imposed upon A in Dowd et al. (2011b).

Equation 7.6 is not well-identified. In practice, this has the consequence that the gravity model can be difficult to fit to historical time series and may give implausible values.

We might conjecture that a solution to this problem would be to allow for deterministic trends up to linear order in the cointegrating relationship, i.e., using $\boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t$ in place of $\boldsymbol{\nu}$ in Equation 7.6 to give

$$\Delta \begin{pmatrix} \kappa_t^{(I)} \\ \kappa_t^{(II)} \end{pmatrix} = \begin{pmatrix} \nu_0^{(I)} \\ \nu_0^{(II)} \end{pmatrix} + \begin{pmatrix} \nu_1^{(I)} \\ \nu_1^{(II)} \end{pmatrix} t + \begin{pmatrix} 0 \\ \phi \end{pmatrix} \begin{pmatrix} 1, & -1 \end{pmatrix} \begin{pmatrix} \kappa_{t-1}^{(I)} \\ \kappa_{t-1}^{(II)} \end{pmatrix} + \begin{pmatrix} \epsilon_t^{(I)} \\ \epsilon_t^{(II)} \end{pmatrix}$$

$$\Delta \boldsymbol{\kappa}_t = \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \alpha \beta^\top \boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \tag{7.10}$$

Such a model is well-identified as it does not change form under the transformation in Equation 7.9

$$\begin{aligned} \Delta \hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\nu}_0 + \boldsymbol{\nu}_1 t + \boldsymbol{c} - \alpha \beta^\top (\boldsymbol{a} + \boldsymbol{c} t) + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \hat{\boldsymbol{\nu}}_0 + \hat{\boldsymbol{\nu}}_1 t + \alpha \beta^\top \hat{\boldsymbol{\kappa}}_{t-1} + \boldsymbol{\epsilon}_t \\ \hat{\boldsymbol{\nu}}_0 &= \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha \beta^\top \boldsymbol{a} \\ \hat{\boldsymbol{\nu}}_1 &= \boldsymbol{\nu}_1 - \alpha \beta^\top \boldsymbol{c} \end{aligned}$$

However, because we have the first difference of the time series on the left-hand side of Equation 7.10, when we integrate this equation, we obtain quadratic trends in the levels of the period functions. This is undesirable as we do not generally observe quadratic trends in the fitted parameters and they might change direction when projected into the future with near certainty for no compelling biological reason. Therefore, the model in Equation 7.10 conflicts with our desire for biologically reasonable[7] projections.

There is, however, a way to obtain both biological reasonableness and identifiability under the transformations in Equation 7.4. This is to restrict the linear deterministic trend in Equation 7.10 by imposing $\boldsymbol{\nu}_1 = \alpha \beta_1$ where β_1 is an arbitrary constant. This will ensure that the relevant deterministic trend is present in $\boldsymbol{\kappa}_t$, but is constrained within the stationary cointegrating relationships and is not present in the non-stationary part of the relationship.

[7] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*"

This means that we need to include constrained deterministic linear trends in the cointegrating relationship, but leave an unconstrained constant term, i.e.

$$\begin{aligned}\Delta\boldsymbol{\kappa}_t &= \boldsymbol{\nu}_0 + \alpha\beta_1 t + \alpha\beta^\top\boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\nu}_0 + \alpha(\beta^\top\boldsymbol{\kappa}_{t-1} + \beta_1 t) + \boldsymbol{\epsilon}_t\end{aligned} \tag{7.11}$$

To see that this structure is well-identified under the transformations in Equation 7.4, let us transform the parameters using Equation 7.9 to obtain

$$\begin{aligned}\Delta\hat{\boldsymbol{\kappa}}_t &= \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha\beta^\top\boldsymbol{a} + \alpha\left(\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + (\beta_1 - \beta^\top\boldsymbol{c})t\right) + \boldsymbol{\epsilon}_t \\ &= \hat{\boldsymbol{\nu}}_0 + \alpha\left(\beta^\top\hat{\boldsymbol{\kappa}}_{t-1} + \hat{\beta}_1 t\right) + \boldsymbol{\epsilon}_t\end{aligned}$$

where $\hat{\boldsymbol{\nu}}_0 = \boldsymbol{\nu}_0 + \boldsymbol{c} - \alpha\beta^\top\boldsymbol{a}$, as previously, and $\hat{\beta}_1 = \beta_1 - \beta^\top\boldsymbol{c}$. This model also gives biologically reasonable values for ϕ which do not depend upon the identifiability constraints imposed when fitting the models, as demonstrated in Section 7.4.3.

7.4.2 Cohort parameters

As with the period parameters, the values of $\gamma_y^{(I)}$ and $\gamma_y^{(II)}$ are not uniquely identifiable in the classic APC model, but instead depend upon our choice of identifiability constraints. Equations 7.3 and 7.4 give us the freedom to add linear trends in time to either or both time series independently, i.e.

$$\begin{aligned}\begin{pmatrix}\hat{\gamma}_y^{(II)} \\ \hat{\gamma}_y^{(II)}\end{pmatrix} &= \begin{pmatrix}\gamma_y^{(I)} \\ \gamma_y^{(II)}\end{pmatrix} + \begin{pmatrix}b^{(I)} \\ b^{(II)}\end{pmatrix} + \begin{pmatrix}c^{(I)} \\ c^{(II)}\end{pmatrix} y \\ \hat{\boldsymbol{\gamma}}_y &= \boldsymbol{\gamma}_y + \boldsymbol{b} + \boldsymbol{c}y\end{aligned} \tag{7.12}$$

Rewriting Equation 7.8 in the form

$$\Delta\boldsymbol{\gamma}_y = \boldsymbol{\mu} + A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\beta^\top\boldsymbol{\gamma}_{y-1} + \boldsymbol{\varepsilon}_y$$

we see that this is also not well-identified as it changes form under the transformation in Equation 7.12

$$\Delta\hat{\boldsymbol{\gamma}}_y = \boldsymbol{\mu} + \boldsymbol{c} - A\boldsymbol{c} - \alpha\beta^\top(\boldsymbol{b} + \boldsymbol{c}y) + A\Delta\hat{\boldsymbol{\gamma}}_{y-1} + \alpha\beta^\top\hat{\boldsymbol{\gamma}}_{y-1} + \boldsymbol{\varepsilon}_y$$

as the transformed drift term, $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{c} - A\boldsymbol{c} - \alpha\beta^\top(\boldsymbol{b} + \boldsymbol{c}y)$, is now a linear function in year of birth, y.

However, in the same manner as used for the period parameters above, we can introduce a constrained linear trend into the cointegrating relationship in order to give well-identified projections which are biologically reasonable

$$\Delta\boldsymbol{\gamma}_y = \boldsymbol{\mu} + A\Delta\boldsymbol{\gamma}_{y-1} + \alpha\left(\beta^\top\boldsymbol{\gamma}_{y-1} + \tilde{\beta}_1 y\right) + \boldsymbol{\varepsilon}_y \tag{7.13}$$

This can be shown to be well-identified by transforming the cohort parameters in a similar fashion.

7.4.3 Application to England & Wales and CMI Assured Lives data

In order to illustrate how the original gravity model gives projections of mortality which depend upon the identifiability constraints chosen, we apply the gravity model to the same data used in Dowd et al. (2011b), i.e., the dominant population is the combined populations of England & Wales and the subordinate population is that of assured lives in the UK as recorded by the Continuous Mortality Investigation, i.e., those people who purchase life assurance policies with UK insurance companies. In both cases, we use data for ages 50 to 90 and years 1947 to 2006.[8]

We start by fitting the classic APC model to the data.[9] In doing so, we have a choice over the identifiability constraints imposed on the models for England & Wales and the CMI Assured Lives. We investigate four different sets of identifiability constraints, which were used for the classic APC model in Chapter 4, i.e.,

Case 1: $\sum_t \kappa_t = 0$, $\sum_y n_y\gamma_y = \sum_{x,t}\gamma_{t-x} = 0$ and $\sum_y n_y\gamma_y(y-\bar{y}) = \sum_{x,t}\gamma_{t-x}((t-\bar{t}) - (x-\bar{x})) = 0$.

Case 2: $\sum_t \kappa_t = 0$, $\sum_y \gamma_y = 0$ and $\sum_y \gamma_y(y-\bar{y}) = 0$.

[8] Data for England & Wales is taken from Human Mortality Database (2014) and we are indebted to the Continuous Mortality Investigation for providing CMI Assured Lives dataset.

[9] To do this, we use a two-step procedure to fit the model for simplicity, i.e., we fit the classic APC model to the data first, and then fit the time series process to the fitted parameters using a least squares approach. This is in contrast with the approach used in Dowd et al. (2011b), where a one-step method is used. However, Dowd et al. (2011b) introduce additional parameters into the one-step method in a Bayesian-type approach whose purpose appears to be to constrain the value of ϕ and prevent it taking values which are not biologically reasonable. However, as discussed later, this issue arises because the gravity model is not well-identified and therefore should not be necessary in a well specified model.

Case 3: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(x - \bar{x}) = 0$.

Case 4: $\sum_t \kappa_t = 0$, $\sum_{x,t} \gamma_{t-x} = 0$ and $\sum_{x,t} \gamma_{t-x}(t - \bar{t}) = 0$.

We investigate the constraints shown in Case 1 and Case 3 as they are the "natural" constraints and the constraints used in Dowd et al. (2011b), respectively, as discussed in Section 7.2. The constraints in Case 2 are similar to those in Case 1, except that the summations are taken over each year of birth rather than over all ages and years in the dataset. This has the effect of moving from a weighted average of the cohort parameters being equal to zero (with the weights determined by the number of observations for each cohort) in Case 1 to a simple arithmetical average in Case 2, and similarly for the linear trend. Although not used for the classic APC model, similar constraints were imposed on the cohort term in Model M6 in Cairns et al. (2009) and so have been included for comparison. As discussed in Section 7.2, the logic underpinning the selection of the Case 3 constraints in Dowd et al. (2011b) was that the static age function in the model should explain all the observed linearity across ages. We can apply similar logic to the period function in the classic APC model, i.e., that the period function, κ_t should explain all of the observed linearity across time, to give the constraints in Case 4.

It is important to note that all four sets of constraints were developed to give the same demographic significance to the cohort parameters, i.e., that they should be centred around zero and the other functions in the model should capture any linear trends. Because of this, these four sets of constraints give very similar sets of fitted parameters when these are plotted. These sets of parameters also give identical fitted mortality rates, since they can be transformed into each other using Equations 7.2, 7.3 and 7.4. However, the different sets of parameters are not identical. We therefore see that demographic significance, whilst helpful in selecting an appropriate set of identifiability constraints, does not specify a single, unique set of constraints to use. Model users with the same interpretation of the parameters can reasonably choose to impose different constraints and obtain different fitted parameters when using the same model with the same data. Furthermore, the fact that demographic significance is subjective and, in practice, different model users adopt a range of interpretations for the different parameters highlights the fact that we must take care to ensure that the projected mortality rates are independent of the arbitrary choice of constraints made when fitting the model, and underscores the extent to which the identifiability constraints we choose is arbitrary.

In each case, we apply the same identifiability constraints to both populations. Figure 7.1 shows the fitted values of $\kappa_t^{(I)} - \kappa_t^{(II)}$ using the Case 1 constraints. Dowd et al. (2011b)

assumed that these differences are stationary, however, Figure 7.1 shows that they have a clear linear trend which would bias the estimation of ϕ in the original specification of the model. Since the magnitude and direction of this trend is dependent upon the identifiability constraints imposed, the degree of this bias is dependent upon our choice of identifiability constraints. However, the modified gravity model allows for the potential presence of a linear trend in the cointegrating relationship and therefore any estimates for ϕ will not be biased by such a trend.

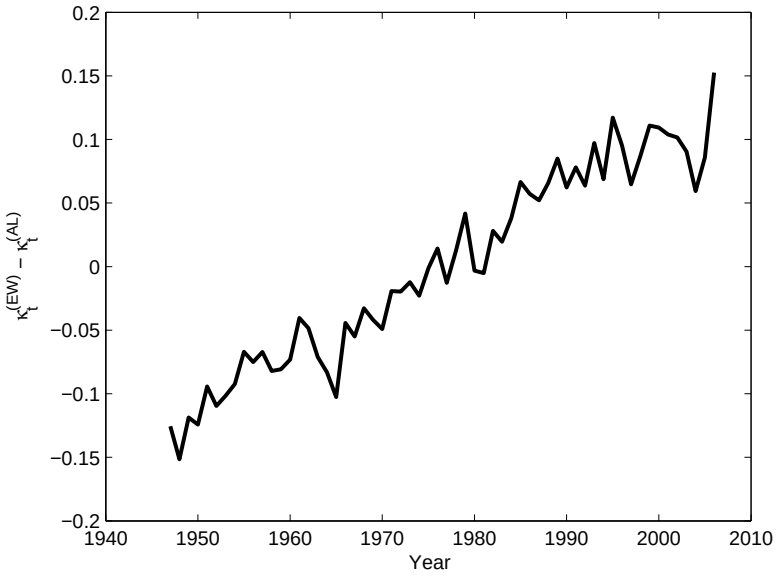


FIGURE 7.1: Difference between the period functions

To demonstrate this numerically, for each set of fitted period parameters, we then first fit the original gravity model in Equation 7.6 and then the modified model in Equation 7.11. We pay particular attention to the estimated value of ϕ found, as this will determine the rate at which divergence between the two populations mean reverts.

	Original gravity model	Modified gravity model
Case 1	0.0706	0.3234
Case 2	0.0702	0.3234
Case 3	0.0701	0.3234
Case 4	0.0700	0.3234

TABLE 7.1: Values of ϕ for different identifiability constraints

The results shown in Table 7.1 indicate that the rate of mean reversion (and therefore the distribution of projected mortality rates) is dependent upon the identifiability constraints using the original gravity model, whereas this is not the case for the modified model. The differences between the cases for the original model appear relatively small.

However, this is because the four sets of identifiability constraints used were selected on the basis of the same demographic significance for the parameters and therefore the fitted parameters were broadly comparable. This will not necessarily always be the case, as demographic significance is subjective and different model users may have very different understandings as to the interpretation of the parameters.

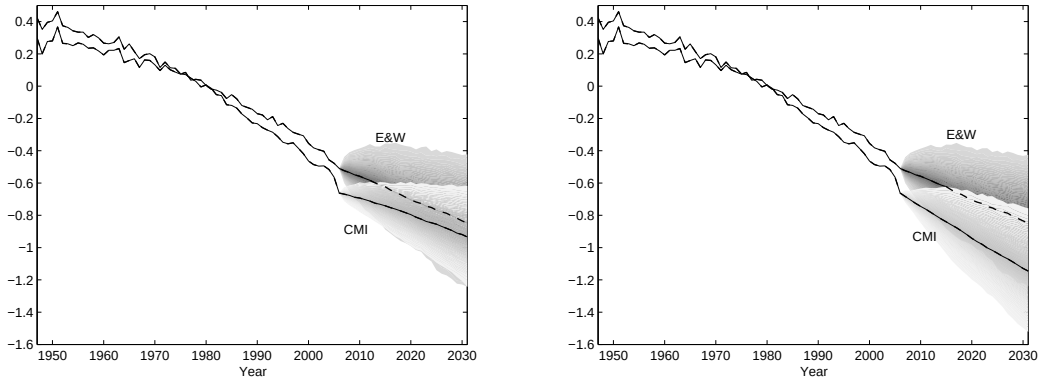
The most important point is not how small the differences are but that they are different at all. The identifiability constraints made no difference to the the fitted mortality rates for the different cases - they were *identical*. However, the distribution of the projected mortality rates depends upon ϕ, which varies between the four cases in the original specification of the model. Therefore, the projected mortality rates would depend upon the choice of identifiability constraints. This is inconsistent with the fitting stage, where the choice of identifiability constraints made no difference to the fitted mortality rates. By contrast, the modified gravity model avoids this, as shown by the fitted value of ϕ being identical in all four cases in Table 7.1.

In particular, we note that it is possible that some sets of identifiability constraints for the classic APC model would give values of ϕ in the original gravity model which were greater than unity or less than zero. Therefore, the arbitrary choice of identifiability constraint may lead to diverging projections of mortality in the original gravity model, despite having the same historical fitted mortality rates as the cases shown. This is clearly something which should be avoided by use of the modified gravity model.

It is also interesting to note that the modified gravity model gives values for ϕ which are considerably larger than in the original model. This is because the parameter now captures the genuine reversion between the period functions (i.e., the saw-tooth pattern in Figure 7.1) without additionally trying to capture the linear trend.

These modifications make a significant difference to the projected parameters when using the gravity model, as shown in Figure 7.2 using the Case 1 identifiability constraints.

As can be seen in Figure 7.2a, the original specification of the gravity model adjusts the central trend so that there is a sharp change of trend in the CMI population at the point where the projections begin. In contrast, the modified gravity model in Figure 7.2b allows the trends observed in either population to continue in future.



(A) Original gravity model (B) Modified gravity model

FIGURE 7.2: Projected period parameters

The change in trend exhibited by the period parameters in the original gravity model is explained by the transition between the past, where the linear trends are diverging in the period parameters fitted to the historical data, and the future, where the gravity model is forcing them together. Since the linear trends in the fitted period parameters were unidentifiable and, hence, entirely dependent upon the identifiability constraints imposed upon the model, the magnitude of the trend change also depends solely upon the arbitrary identifiability constraints. Therefore, the existence of such a trend change is not well-identified and leads to projected mortality rates which depend upon the identifiability constraints chosen, unlike the fitted mortality rates.

Furthermore, the existence of such a trend change leads to inconsistencies between the past and the future. This is not compatible with the extrapolative approach to projecting mortality, as discussed in Chapter 6. Although there might be insufficient evidence in the historical data to support the existence of changes in trend in the fitted period parameters, the original gravity model imposes a trend change, precisely at the transition between the historical data and the projected mortality rates. One implication of this is that the data has been collected at a unique point in time that is qualitatively different from the periods before or after it. We do not believe that such an assumption is tenable.

In contrast, the modified gravity model does not predict a change in trend at the transition between past and future. As discussed in Chapter 4, the linear trends in the classic APC model are unidentifiable and depend entirely upon the identifiability constraints, whereas the variation around those trends is identifiable. Therefore, the modified gravity model leaves the linear trend in both populations unchanged, but allows the variation around these trends to be cointegrated. This means that decreases in mortality which

are faster than expected in England & Wales are correlated with faster than expected declines in mortality rates in the CMI Assured Lives population. Capturing this correlation is vital in the measuring of basis risk between populations, as in Li and Hardy (2011) and Coughlan et al. (2011), and when modelling liabilities and securities which depend upon mortality in multiple populations, as discussed in Chapter 8.

Not only is the modified gravity model well-identified, we also believe that it gives projections which give greater consistency between the past and the future. The behaviour of the fitted parameters has been analysed and projected into the future, without assuming a priori that this behaviour will change. Such an approach is far more consistent with the extrapolative approach to projecting mortality rates discussed in Section 6.2 of Chapter 6 than the assumption of a trend change present in the original gravity model.

Furthermore, we believe that an assumption whereby projections maintain the same trends in each population but allow for correlated variation around these trends is more justified in terms of biological reasonableness than assuming that the period parameters converge in future. The factors impacting deviations in mortality rates from trend in one population are likely to be common across populations, leading to correlated variation around the trend in the two populations. In contrast, the differing trend rates of mortality improvement are likely to be generated by more fundamental socio-economic causes, which will remain unchanged for the foreseeable future.

In summary, we find that the modified gravity model gives projected mortality rates for England & Wales and the CMI Assured Lives populations which are well-identified and have variation which is correlated in a biologically reasonable fashion. However, the modified gravity model does not induce the trends present in either population to change sharply at the transition point between past and future, which is a feature of the original gravity model and which was imposed to ensure that mortality rates in the two populations are "coherent".

7.4.4 Coherence

The term "coherence" was introduced in Li and Lee (2005), and was defined formally in Hyndman et al. (2013) in terms of the relative mortality rates between populations, i.e.,

$$\mathbb{E}\left[\frac{\mu_{x,t}^{(p_1)}}{\mu_{x,t}^{(p_2)}}\right] \rightarrow R_x \tag{7.14}$$

a function of age only. This means that relative mortality rates are stationary, and so the mortality rates projected in the two populations do not diverge with time. Coherence is a stronger requirement for a multi-population mortality model than simply allowing the covariation observed in the past to continue into the future, as discussed in Section 7.4.3 above.[10]

The original gravity model was introduced in part to ensure that mortality rates in the England & Wales and CMI Assured Lives populations are coherent. The original gravity model has coherence built into it, since

$$\begin{aligned}\ln\left[\frac{\mu_{x,t}^{(I)}}{\mu_{x,t}^{(II)}}\right] &= \left(\alpha_x^{(I)} - \alpha_x^{(II)}\right) + \left(\kappa_t^{(I)} - \kappa_t^{(II)}\right) + \left(\gamma_{t-x}^{(I)} - \gamma_{t-x}^{(II)}\right) \\ &= \left(\alpha_x^{(I)} - \alpha_x^{(II)}\right) + \beta^{\top}\left(\boldsymbol{\kappa}_t + \boldsymbol{\gamma}_{t-x}\right)\end{aligned}$$

which is stationary in time by construction.[11]

However, when the gravity model is modified to ensure projections are well-identified, coherence no longer necessarily holds, since we can have different linear trends in both populations (i.e., $\beta^{\top}\left(\boldsymbol{\kappa}_t + \beta_1 t + \boldsymbol{\gamma}_{t-x} + \tilde{\beta}_1(t-x)\right)$ is stationary, whilst $\beta^{\top}\left(\boldsymbol{\kappa}_t + \boldsymbol{\gamma}_{t-x}\right)$ is not). The level of divergence will be set by the observed divergence between the populations in the historical dataset, i.e., we will project mortality rates that will continue to diverge if they have been observed to do so in the past. Such an approach gives greater consistency between the historical data and projected mortality rates.

Therefore, we see that there is the potential for conflict between the desire for coherent projections and the need for projections of the model to be well-identified. In general, we believe that obtaining projected mortality rates that do not depend on arbitrary choices made when fitting the model to data is more important than a desire to prevent divergence between populations, for the reasons discussed below. However, we note that identifiability issues in mortality models are features of the *parameters* in mortality models, whereas coherence is a property of the projected mortality *rates*, which should be

[10] Coherence is a potential feature of the projected mortality rates and can result from a number of different techniques for projecting mortality, rather than it being a technique in itself. For instance, the original and modified gravity models both involve the technique of cointegration, but one gives coherent projected mortality rates, whilst the other does not. Conversely, the original gravity model and the SAINT model of Jarner and Kryger (2011) both give coherent mortality rates, but use different techniques to achieve this.

[11] However, the long-run distribution of $\frac{\mu_{x,t}^{(I)}}{\mu_{x,t}^{(II)}}$, and specifically R_x, will depend upon the arbitrary identifiability constraints imposed when fitting the model.

independent of these issues. If coherence is desired, we therefore believe that methods of imposing it should focus on constraining the projected mortality rates themselves, rather than specific features of the model parameters, which will depend on the identifiability constraints imposed.

However, we would often go further and question the desire to impose coherence a priori on projected mortality rates. Much of the work discussing coherent projections of mortality rates has been based on the idea that mortality rates should not diverge indefinitely in future between related populations. For instance, Li and Lee (2005) stated that *"Obviously, mortality differences between [closely related] populations should not increase over time indefinitely if the similar socio-economic conditions and close connections were to continue."* We believe that there are two problems with this conjecture.

First, whilst it might be true that projecting divergences indefinitely into the future may be unrealistic, we would point out that extrapolating any model indefinitely into the future is likely to give nonsensical results sooner or later. For example, the Lee-Carter model will tend to give mortality rates arbitrarily close to zero at all ages if projected far enough into the future. However, such a phenomenon is more the fault of a modeller misusing the model to make inappropriate forecasts than it is the fault of the model itself. A general rule of thumb is that a model should not be projected for a longer period than the data used to estimate it. Given this, the question becomes why we should believe that mortality differences cannot diverge for another 50 years (say) if we have observed mortality differences diverging for the previous 50 years. Assuming that the evolution of mortality rates in the future will be qualitatively different from the past is inconsistent with the extrapolative approach.

Second, we believe that it is simply untrue that differences in mortality rates cannot persist for prolonged periods between ostensibly related populations. For example, life expectancy at age 65 varies considerably between areas in the same city[12] in a pattern which has been stable for decades, let alone between different socio-economic groups within the same country (see Harper et al. (2007) and Villegas and Haberman (2014)) or between countries. Whilst coherence does not impose the requirement that these long-established differences decrease, it does assume that they are not expected to grow beyond their current level, which we do not believe is supported by the evidence. It also raises the question as to what is so special about the currently observed differences in mortality that they should act as a barrier beyond which further divergence is not

[12]Source: http://data.london.gov.uk/dataset/life-expectancy-birth-and-age-65-ward

possible.

Therefore, we do not believe that coherence is a desirable property to impose upon an extrapolative multi-population mortality model. As scientific investigators, we should allow the data to speak for itself rather than impose any prior views onto the models that we use. This is consistent with the extrapolative approach discussed in Section 6.2 of Chapter 6, where analysis of historical data, rather than subjective opinions and biases, is used to project mortality rates. If the data supports our beliefs, that is encouraging. If the data does not, then we need to examine either our preconceptions to determine whether they need to be revised or re-examine the model we are using to analyse and project the data.

Ultimately, many of the preconceptions which lead to a desire for coherence between different populations have a basis in our knowledge of the specific populations under consideration and the specific factors causing the divergence in these populations. For example, the observed divergence between mortality rates in the England & Wales and CMI Assured Lives populations could be attributed to the selective nature of the CMI Assured Lives dataset, which consists of individuals who are likely to be wealthier than the average citizen of England & Wales. In addition, this selective population may adopt different lifestyles, with less smoking and a better diet than the wider population, for example, leading to a differing pattern of mortality. We might reasonably feel that such differences will get less important with time and the wider population adopts the same lifestyle as the sub-population, and therefore that mortality rates in the two population should stop diverging in future.

However, this kind of argument for imposing coherence on a model makes use of additional information regarding the causes of any divergence, information that was not used when fitting the model. We therefore believe that, rather than imposing coherence on a model to obtain the results we want, it would be better to incorporate into our model the additional information that justifies our desire for coherence in the first place. Such information may include economic and lifestyle variables, for instance, as in Reichmuth and Sarferaz (2008), Wang and Preston (2009) and French (2014). This may help explain any observed divergence in the past and potentially allow for coherent projections which are still well founded in a rigorous analysis and extrapolation of the data.

7.5 Identifiability in the cointegrated Lee-Carter model

Zhou et al. (2014) applied a similar cointegration framework as developed for the gravity model to the period parameters of the Lee-Carter model

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t \tag{7.15}$$

for multiple populations. The period parameters are projected using a time series process of the form

$$\Delta\boldsymbol{\kappa}_t = \boldsymbol{\nu} + \Gamma\Delta\boldsymbol{\kappa}_{t-1} + \alpha\beta^\top\boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \tag{7.16}$$

which is a cointegrated relationship of the form in Equation 7.7.[13] As in Dowd et al. (2011b), β was constrained so that $\beta = \begin{pmatrix}1, & -1\end{pmatrix}^\top$ in order that relative mortality rates do not diverge in the two populations. However, no assumption is made regarding the dominance of one population over the other, and therefore no constraint is made on α, unlike the gravity model where $\alpha = \begin{pmatrix}0, & \phi\end{pmatrix}^\top$ was used to impose the condition that population I dominates population II.

As discussed in Lee and Carter (1992) and Chapter 3, the Lee-Carter model is also not well-identified and possesses the invariant transformations

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \{\alpha_x, \frac{1}{a}\beta_x, a\kappa_t\} \tag{7.17}$$

$$\{\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t\} = \{\alpha_x - b\beta_x, \beta_x, \kappa_t + b\} \tag{7.18}$$

which are used to impose identifiability constraints in a similar fashion to the classic APC model. These invariant transformations can be applied independently to the two populations without affecting the fitted mortality rates, and so we can write

$$\hat{\boldsymbol{\kappa}}_t = A\left(\boldsymbol{\kappa}_t + \boldsymbol{b}\right) \tag{7.19}$$

with $A = \begin{pmatrix} a^{(I)} & 0 \\ 0 & a^{(II)} \end{pmatrix}$.

[13] Again, the form of Equation 7.16 differs from the form of Equation 7.7 due to the stationary cointegrating term, $\alpha\beta^\top\boldsymbol{\kappa}_{t-1}$, as opposed to $\alpha\beta^\top\boldsymbol{\kappa}_{t-2}$ required by Equation 7.7. However, this can be resolved in the manner outlined in footnote 6.

If we apply this transformation to the time series process in Equation 7.16 we obtain

$$\begin{aligned}\Delta\hat{\boldsymbol{\kappa}}_t &= A\boldsymbol{\nu} - A\alpha\beta^\top \boldsymbol{b} + A\Gamma A^{-1}\Delta\hat{\boldsymbol{\kappa}}_{t-1} + A\alpha\beta^\top A^{-1}\hat{\boldsymbol{\kappa}}_{t-1} + A\boldsymbol{\epsilon}_t \\ &= \hat{\boldsymbol{\nu}} + \hat{\Gamma}\Delta\hat{\boldsymbol{\kappa}}_{t-1} + \hat{\alpha}\hat{\beta}^\top\hat{\boldsymbol{\kappa}}_{t-1} + \hat{\boldsymbol{\epsilon}}_t\end{aligned}$$

which is of the same form as Equation 7.16 if we redefine the terms appropriately. In particular, this involves setting

$$\begin{aligned}\hat{\beta} &= A^{-1}\beta \\ &= \begin{pmatrix} \frac{1}{a^{(I)}} & 0 \\ 0 & \frac{1}{a^{(II)}} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= \left(\frac{1}{a^{(I)}}, \quad -\frac{1}{a^{(II)}}\right)^\top\end{aligned}$$

i.e., if the time series process is well-identified, β cannot be restricted to have any particular form, since these restrictions will only apply for one set of identifiability constraints. We also see that we are free to set $\hat{\alpha} = A\alpha$, since α is not constrained to any particular form initially. Therefore, in order for the model of Zhou et al. (2014) to be well-identified, the restriction on β as well as the restriction on α must also be relaxed. This was commented upon in Nielsen and Nielsen (2014).

The reason for the difference between the models of Zhou et al. (2014) and Dowd et al. (2011b) arises because of the differences in the underlying APC mortality models used in either study. In the Lee-Carter model used in Zhou et al. (2014), the "scale" of the period functions is defined by an identifiability constraint on β_x. This scale is arbitrary, and we can change it without affecting the fitted mortality rates from the model. Therefore the projected mortality rates from the model of Zhou et al. (2014) also need to also be invariant to changes in this scale. In contrast, the scale of the period functions is defined by the parametric age function in the classic APC function, and not by an identifiability constraint. Therefore, it cannot be changed in the model, and so we do not have to ensure that the projected mortality rates are invariant to changes in its scale.

Conversely, the Lee-Carter model does not have unidentifiable linear trends, unlike the classic APC model. Therefore the model of Zhou et al. (2014) does not require a linear drift term, i.e., $\beta_1 t$, in the cointegrating relationships. The classic APC model does contain unidentifiable linear trends in the parameters, which can be varied in the historic parameters without affecting the fitted mortality rates. It is, therefore, essential that the projected mortality rates from the model of Dowd et al. (2011b) are invariant to changes in the linear trend in the period an cohort functions, which is ensured by the presence

of the linear drift term, $\beta_1 t$, in the cointegrating relationships.

We see, therefore, that the form the gravity model needs to take in order to be well-identified depends on the underlying APC mortality model being used and the identifiability issues within that particular model. It is therefore essential that these identifiability issues are fully analysed and understood, as discussed in Chapters 3 and 4. In general, we see that it is best to avoid making any impositions on the structure of α and β, and so use the most general form of cointegrating relationship, in order to avoid any potential identifiability issues and avoid constraining the form of the model unnecessarily and, potentially, inappropriately.

7.6 Extending the cointegration model

For the general cointegration model in Equation 7.7, our approach generalises naturally to models where there are unidentifiable higher-order polynomial deterministic trends in the parameters. If the period functions of a model have unidentified deterministic trends which are polynomial of order M, then in order to be well-identified under the corresponding invariant transformations, we will need to allow for unconstrained deterministic trends up to polynomial order $M-1$ and constrained deterministic trends of order M.

For instance, the model of Plat (2009a)

$$\ln(\mu_{x,t}^{(p)}) = \alpha_x^{(p)} + \kappa_t^{(1,p)} + (x-\bar{x})\kappa_t^{(2,p)} + (x-\bar{x})^+\kappa_t^{(3,p)} + \gamma_{t-x}^{(p)} \tag{7.20}$$

has unidentifiable quadratic trends, as discussed in Chapter 4. If the Plat (2009a) model were fitted to two populations, we would have six period functions in total - $\kappa_t^{(1,I)}$ and $\kappa_t^{(1,II)}$ with unidentified quadratic trends, $\kappa_t^{(2,I)}$ and $\kappa_t^{(2,II)}$ with unidentified linear trends and $\kappa_t^{(3,I)}$ and $\kappa_t^{(3,II)}$ with unidentified constants.

We could look for a cointegration model involving all six period functions. Cointegration, by its nature, involves interactions between the different period functions. We therefore are unable to allow for deterministic trends of different order in different period functions. Allowing for cointegration between all six time series would therefore mean allowing for constrained quadratic trends and unconstrained linear trends in all six period functions,

which may lead to projections which are not biologically reasonable for each population.

It is more biologically reasonable to consider each pair of period functions separately based on their shared demographic significance. This would mean looking for cointegrating relationships with constrained quadratic (and unconstrained constant and linear) trends for the two $\kappa_t^{(1,p)}$ functions, relationships with constrained linear trends for the $\kappa_t^{(2,p)}$ functions, and so on. That is, we use

$$\Delta\boldsymbol{\kappa}_t^{(1)} = \boldsymbol{\nu}_0^{(1)} + \boldsymbol{\nu}_1^{(1)} t + \alpha^{(1)} \left(\beta^{(1)\top} \boldsymbol{\kappa}_{t-1}^{(1)} + \beta_2^{(1)} t^2 \right) + \boldsymbol{\epsilon}_t^{(1)} \tag{7.21}$$

$$\Delta\boldsymbol{\kappa}_t^{(2)} = \boldsymbol{\nu}_0^{(2)} + \alpha^{(2)} \left(\beta^{(2)\top} \boldsymbol{\kappa}_{t-1}^{(2)} + \beta_1^{(2)} t \right) + \boldsymbol{\epsilon}_t^{(2)} \tag{7.22}$$

$$\Delta\boldsymbol{\kappa}_t^{(3)} = \alpha^{(3)} \left(\beta^{(3)\top} \boldsymbol{\kappa}_{t-1}^{(3)} + \beta_0^{(3)} \right) + \boldsymbol{\epsilon}_t^{(3)} \tag{7.23}$$

to project the period functions. This approach is used in Chapter 8, albeit in a model with unidentified cubic (as opposed to merely quadratic) trends.

7.7 Conclusions

Cointegration can be a powerful tool for projecting mortality rates in related populations. However, it is a tool which must be used with care to ensure that we have identifiability under any invariant transformations which allocate unidentifiable polynomial trends between the parameters. In the case of the gravity model of Dowd et al. (2011b) and the model of Zhou et al. (2014), we have shown how to adapt the process used to project the period functions so that it gives well-identified projections that do not depend on the arbitrary identifiability constraints imposed. We have also shown how this can be generalised to more complicated APC mortality models.

Further, we have shown that we cannot also impose the condition that mortality rates are coherent and do not diverge in future. Not only does imposing coherence mean that the projected mortality rates will depend upon the arbitrary identifiability constraints selected, it is also incompatible with an extrapolative approach to modelling mortality. An extrapolative approach must, first and foremost, take its lead from the evidence of the historical data. While, in many circumstances, a belief in coherence is quite natural, we believe we should test for its existence in the historical data statistically using well-identified models, rather than assume its existence beforehand as an article of faith. If we do not find any evidence for coherence in the historical data, this should be considered a puzzle to explain using more data and better models, and not just an error to be corrected by an ad hoc fix which overrides the evidence of the data to obtain the

results we anticipated in advance. Such an approach is not only more rigorous and more scientific, but can also give new insights into the factors which govern the evolution of mortality rates and enhance our understanding of longevity risk.

Chapter 8

Modelling Longevity Bonds: Analysing the Swiss Re Kortis Bond

8.1 Introduction

The traded market for longevity risk continues to grow and develop. As the risks posed by increasing longevity for the providers of pensions and annuities have gained greater prominence, a variety of different vehicles have been proposed and implemented to transfer longevity risk to the capital markets. These have included bonds, swaps and forwards, each linked to different measures of mortality rates and survivorship.

A key contribution to the development of the market was the issuance of the Kortis bond by Swiss Re in 2010. Unlike previous mortality and longevity securitisations, the Kortis bond is linked to the divergence in mortality improvement rates between two countries, rather than to mortality rates directly or to survivorship amongst a cohort. As such, it was promoted as the first "longevity trend bond". The bond might herald a distinctly new way of transferring the risk of faster than expected reductions in mortality rates, from insurers and reinsurers to investors willing to hold these risks as part of a diversified portfolio.

The development of new longevity-linked securities has been aided by and, in turn, encouraged the development of increasingly sophisticated mortality models. These are necessary in order to estimate accurately the risk present in such securities. In particular, they need to project mortality rates with complex correlation structures, robustly estimated cohort effects and dependencies between different populations. Such projections

Part III

Modelling Mortality for Pension Schemes

Chapter 9

Basis Risk and Pension Schemes: A Relative Modelling Approach

9.1 Introduction

Longevity risk is increasingly recognised as a major risk in developed countries, as rising life expectancies place unanticipated strains on social security and healthcare systems (see Oppers et al. (2012)). As well as being of concern for governments, however, longevity risk also affects private organisations that have promised people an income for life, be this in the form of an insured annuity or an occupational pension. In the UK, this means that longevity risk affects the thousands of occupational pension schemes[1] established by companies during the 20^{th} century to provide final salary pensions to their employees.

However, when it comes to managing the longevity risk in a pension scheme, actuaries face a critical problem: a shortage of mortality data for the scheme. A typical UK pension scheme has fewer than 1,000 members and may have reliable, computerised member records going back no more than a decade. This is insufficient for use with the sophisticated stochastic mortality models that have been developed in recent years to measure longevity risk in national populations, since these models require more data to estimate parameters robustly and longer time series to make projections into the future. While the insights gained from the study of national populations are useful for the study of longevity risk in pension schemes, actuaries are left with a nagging doubt: "What if my

[1] In this chapter, we refer to "pension schemes" which administer the provision of defined benefits to members. We draw a semantic distinction between a "pension scheme" and a "pension plan", which we would use as a more general term for any defined benefit or defined contribution pension arrangement provided on either a group or an individual basis.

scheme is different from the national population?" The potential for divergence in mortality rates between the scheme and the national population is called "basis risk", and, anecdotally, is often given as a key reason holding back the use of standardised financial instruments (based on national data) to manage longevity risk in pension schemes.

The actuarial profession in the UK initiated the Self-Administered Pension Scheme study in 2002 in an attempt to overcome these issues with data. The study pools data from almost all large occupational pension schemes in the UK, allowing insights about how typical pension schemes differ from the national population to be established.

In this study, we use the data collected by the Self-Administered Pension Scheme study and develop a "relative" model for mortality in order to compare the evolution of mortality rates in UK occupational pension schemes directly with that observed in the national population. Such a relative model has the advantages of parsimony and robustness, important properties when dealing with the smaller datasets available for pension schemes. We then use this relative model to investigate the phenomenon of basis risk between pension schemes and the UK population, as well as the potential of using this approach on even smaller populations comparable with the size of an individual scheme. In doing so, we bring into question the potential importance of basis risk in small populations and find that in most contexts it is likely to be substantially outweighed by other risks in a pension scheme. This is investigated further in Chapter 10 .

The outline of this chapter is as follows. Section 9.2 describes the Self-Administered Pension Schemes (SAPS) study and how the population observed by it differs structurally from the national UK population. Section 9.3 discusses the "relative" modelling framework we will use to compare the mortality experience of these populations. Section 9.4 then applies this framework to data from the SAPS study, tests the models produced and considers the impact of parameter uncertainty on these conclusions. Section 9.5 uses the relative model to project mortality rates for the sub-population in the context of assessing the basis risk between it and the national population. Section 9.6 then assesses the feasibility of using the relative model for smaller populations which have sizes more comparable to those of actual UK pension schemes. Section 9.7 discusses some of the broader conclusions on the importance of basis risk we draw from this study, whilst Section 9.8 summarises our findings.

9.2 The Self-Administered Pension Scheme study

The Institute of Actuaries in England & Wales and the Faculty of Actuaries in Scotland initiated the SAPS study in 2002 to investigate the mortality experience of pensioner members of occupational pension schemes in the UK. Data from the SAPS study has been analysed by the Continuous Mortality Investigation (CMI) to produce the graduated mortality tables[2] in use by the majority of pension schemes in the UK for funding and accounting purposes.[3] The CMI has also analysed the SAPS data in terms of the evolution of mortality during the study period[4] and the differences in experience for schemes whose employers are in different industries.[5]

UK pension schemes with more than 500 pensioner members are asked to submit mortality experience data to the SAPS study after each triennial funding valuation. The CMI provides summaries of the aggregate of this data to members of the study, categorised across a number of different variables, at regular intervals.[6] We have been provided with this data in a more complete form, comprising exposures to risk and death counts (unweighted by the amount of pension in payment) for all men and women in the SAPS study between 2000 and 2011 by the CMI. A summary of the data used in this study is given in Appendix 9.A.

Since it is sampling from a distinct subset of the national population, the dataset collected by the SAPS study is atypical of the UK population data for a number of reasons:

- The dataset is the mortality experience of members of occupational, defined-benefit pension schemes. Typically, this will exclude the unemployed, the self-employed, those employed in the informal sector or those working for newer companies (which typically do not offer defined-benefit pensions).
- The dataset is the mortality experience of members of reasonably large pension schemes. According to The Pensions Regulator (2013b), only around 20% of UK pension schemes have more than 1,000 member in total, a large number of whom are likely to be below retirement age. This means that employees of large, mature companies are likely to be over-represented in the SAPS study.

[2] The S1 tables in Continuous Mortality Investigation (2008) and the S2 tables in Continuous Mortality Investigation (2014a).
[3] The Pensions Regulator (2013a) and Sithole et al. (2012).
[4] See Continuous Mortality Investigation (2011).
[5] See Continuous Mortality Investigation (2012).
[6] See Continuous Mortality Investigation (2014c) for example.

- The dataset is the mortality experience of pension schemes subject to triennial funding valuations. This means that it excludes most public sector employees, who are members of unfunded state pension schemes.
- The dataset is likely to have some individuals in receipt of pensions from multiple sources, for instance, because of employment at two or more different companies, and who will therefore be represented multiple times.
- The dataset will include members of UK pension schemes who emigrate and possibly die overseas, and who therefore would not be included in the UK national population mortality data.

These factors explain why the experience of the SAPS mortality study is believed to be a better proxy for the mortality experience of individual UK pension schemes (even those not included in the SAPS study). The mortality tables graduated from the SAPS data are therefore often used for pension scheme accounting and funding purposes, as opposed to tables graduated from national population data or the experience of individuals buying annuities directly from life insurers. However, they also mean that the future evolution of mortality rates for SAPS members may be different from that of the national population (although they may well be similar in other respects).

Unfortunately, the SAPS dataset poses a number of difficulties for use with the more sophisticated mortality modelling and projection techniques which have been developed in recent years. These include:

- relatively small exposures to risk (at most around 1.5 million members under observation in a single year), leading to greater parameter uncertainty especially in complex models;
- the short length of the study, with only twelve years of data in the sample for analysing the trends present; and
- the method of data collection - schemes submit data in respect of a three-year period at a lag of up to 18 months after the period ends - leads to a distinctive pattern of exposures shown in the data in Appendix 9.A, with only partial data having been submitted to date for the last five years in the study.

For these reasons, it is still advisable to use national mortality data, with its larger exposures and longer period of availability, to produce projections of mortality rates. The SAPS data can then be used to quantify the ways that members of UK pension

schemes are likely to differ from this baseline. We do this by means of a "relative" mortality model, which we now describe.

9.3 Relative mortality modelling

A "relative" mortality model for two populations is one that does not model mortality rates in a smaller population directly, but instead models the relative difference between those rates and those found in a larger, reference population. That is, it models the behaviour of the relative mortality rates, $R_{x,t}$, given by

$$R_{x,t} = f\left(\frac{\mu_{x,t}^{(S)}}{\mu_{x,t}^{(R)}}\right) \tag{9.1}$$

where $\mu_{x,t}^{(p)}$ are the mortality rates in the small population, S, and reference population, R. Typically, mortality rates in the reference population are modelled and projected independently of $R_{x,t}$.

A number of different models of this form have been proposed in order to analyse mortality for various different populations. Those which have explicitly adopted a relative modelling approach include the models of Jarner and Kryger (2011), who used a series of basis functions across age to model $R_{x,t}$ for Denmark compared to the wider EU and assume it mean reverts deterministically in future, and Villegas and Haberman (2014), who investigated the mortality of different socio-economic groups within the UK relative to the national average. However, a good many other multi-population mortality models which have been proposed, such as those of Carter and Lee (1992), Li and Lee (2005), Delwarde et al. (2006), Dowd et al. (2011b), Cairns et al. (2011b), Russolillo et al. (2011) and Wan and Bertschi (2015), can be rewritten as relative mortality models although this was not necessarily commented on by the authors. See Villegas and Haberman (2014) for a useful summary of many of these models and the similarities between them.

The advantage of a relative modelling approach is that it allows us to use a far simpler model for the relative mortality rates, $R_{x,t}$, than would be used for the reference population. This is desirable as we typically have insufficient data for the smaller population to estimate more complex models robustly, but would like to use a sophisticated model for the reference population in order to produce more accurate projections of mortality rates. In addition, there is no requirement that the data for the small population covers

the same range of ages and years as that for the larger population.

9.3.1 The reference model

For the reference population, we choose to use the "general procedure" (GP) of Chapter 5 in order to construct a model sufficient to capture all the significant information present in the national population data. This selects an appropriate model within the class of age/period/cohort (APC) models[7] of the form

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \sum_{i=1}^{N} f^{(R,i)}(x;\theta^{(R,i)})\kappa_t^{(R,i)} + \gamma_{t-x}^{(R)} \tag{9.2}$$

where

- age, x, is in the range $[1, X]$, period, t, is in the range $[1, T]$ and hence that year of birth, y, is in the range $[1 - X, T - 1]$;
- $\alpha_x^{(R)}$ is a static function of age;
- $\kappa_t^{(R,i)}$ are period functions governing the evolution of mortality with time;
- $f^{(R,i)}(x;\theta^{(R,i)})$ are parametric age functions (in the sense of having a specific functional form selected a priori) modulating the impact of the period function dynamics over the age range, potentially with free parameters $\theta^{(R,i)}$;[8] and
- $\gamma_y^{(R)}$ is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as it ages.

The GP selects the number of age/period terms, N, and the form of the age functions $f^{(R,i)}(x)$ in order to construct mortality models which give a close but parsimonious fit to the data. This way, we aim to extract as much information as possible from the national population dataset and have specific terms within the model corresponding to the different age/period or cohort features of interest.

[7] See Chapter 2 for a description of this class of models.

[8] For simplicity, the dependence of the age functions on $\theta^{(R,i)}$ is supressed in notation used in this chapter, although it has been allowed for when fitting the model to data.

9.3.2 The relative model

To analyse the data from the SAPS study, we propose using a model of the form

$$R_{x,t} = \ln\left(\frac{\mu^{(S)}_{x,t}}{\mu^{(R)}_{x,t}}\right) = \alpha^{(\Delta)}_x + \sum_{i=1}^{N} \Lambda^{(i)} f^{(R,i)}(x)\kappa^{(R,i)}_t + \Lambda^{(\gamma)}\gamma^{(R)}_{t-x} + \nu X_{t-x} \tag{9.3}$$

Apart from the νX_y term, this is an APC model of the same form as that used to model the reference population, i.e., with the same age/period terms and cohort parameters. However, these are modulated by the factors $\Lambda^{(j)}$ where $j \in \{1, \ldots, N, \gamma\}$. The νX_{t-x} term, where X_y is a set of deterministic functions of year of birth and ν the corresponding regression coefficients, has been added to the APC structure in order to ensure that the model is identifiable under invariant transformations of the cohort parameters, as discussed in Appendix 9.B.

The choice of structure in Equation 9.3 is also motivated by the fact that we can write the mortality rates for the sub-population as

$$\ln\left(\mu^{(S)}_{x,t}\right) = \alpha^{(S)}_x + \sum_{i=1}^{N} \lambda^{(i)} f^{(R,i)}(x)\kappa^{(R,i)}_t + \lambda^{(\gamma)}\gamma^{(R)}_{t-x} + \nu X_{t-x} \tag{9.4}$$

where $\alpha^{(S)}_x = \alpha^{(R)} + \alpha^{(\Delta)}_x$ and $\lambda^{(j)} = 1 + \Lambda^{(j)}$. We are therefore able to interpret $\alpha^{(\Delta)}_x$ as the difference in the level of mortality between the two populations, whilst the $\lambda^{(j)}$ correspond to the "sensitivity" of the small population to the j^{th} factor in the reference population. In this form, it is possible to see the model as similar in spirit to that proposed by Russolillo et al. (2011), as discussed in Section 9.3.3.

It should be noted that there are two special cases for these sensitivities:

1. $\lambda^{(j)} = 0$ (i.e., $\Lambda^{(j)} = -1$): the small population has no dependence on the j^{th} age/period or cohort term; and
2. $\lambda^{(j)} = 1$ (i.e., $\Lambda^{(j)} = 0$): there is no difference between the reference and small populations with respect to the j^{th} factor.

In order to obtain a more parsimonious model, it may also be desirable to simplify the non-parametric structure[9] for $\alpha^{(\Delta)}_x$ by constraining it to be of a specific parametric form, for example, a linear combination of a set of pre-defined basis functions. However, we

[9] Defined in Chapter 2 as being fitted without any a priori structure or functional form.

must take care when doing so in order that the relative model is robust to changes in the identifiability constraints for the reference model, as discussed in Appendix 9.B.

When fitting the relative model to data, we have a strong preference for parsimony due to the low volume of data for the sub-population. We therefore adopt a "specific-to-general" modelling approach: first testing a highly restricted form of the model with a parametric form for $\alpha_x^{(\Delta)}$ and $\lambda^{(j)} = \{0, 1\}$ and then relaxing these restrictions sequentially. The final model is chosen to maximise the Bayes Information Criteria (BIC),[10] which penalises excessive parameterisation. This procedure is performed algorithmically, and is especially important when we apply the relative model to very small datasets comparable to the size of individual pension schemes, as done in Section 9.6.

9.3.3 Comparison with "three-way Lee-Carter"

It was noted above that many alternative multi-population mortality models have been proposed in the literature, including many which were explicitly designed as relative mortality models and others which can be re-written in relative form. For a summary and comparisons of some of these models, see Li and Hardy (2011) and Villegas and Haberman (2014).

Of these, the model which bears closest resemblance to the model outlined in Section 9.3.2 is the "three-way Lee-Carter" model of Russolillo et al. (2011). This extends the classic model of Lee and Carter (1992) into a third "dimension" of population, beyond the original two dimensions of age and period. They achieve this by including an extra covariate in the Lee-Carter predictor structure to represent the different populations, p, i.e.,

$$\ln(\mu_{x,t}^{(p)}) = \alpha_x^{(p)} + \lambda^{(p)}\beta_x\kappa_t \tag{9.5}$$

The parameters are fitted using multi-dimensional principal components techniques. Villegas and Haberman (2014) pointed out that an additional identifiability constraint is required to obtain a unique set of parameters, which they choose to be $\sum_p \lambda^{(p)} = N_p$, the number of populations. In a two-population setting, this can be re-written as a relative model, with

$$R_{x,t} = \alpha_x^{(\Delta)} + (\lambda^{(S)} - \lambda^{(R)})\beta_x\kappa_t$$

[10] Defined as max(Log-likelihood) $- 0.5 \times$ No. free parameters $\times \ln$(No. data points).

and $\alpha_x^{(\Delta)}$ defined in the same fashion as in Equation 9.3.

We can, therefore, see that the relative model of Section 9.3 can be thought of as a "three-way" extension for multiple populations of the underlying model constructed by the general procedure for a single population, namely

$$\ln\left(\mu_{x,t}^{(p)}\right) = \alpha_x^{(p)} + \sum_i \lambda^{(p,i)} f^{(i)}(x)\kappa_t^{(i)} + \lambda^{(p,\gamma)}\gamma_{t-x}$$

We then introduce the νX_y term in order to ensure that the model does not depend upon the arbitrary identifiability constraints imposed in the reference model, as discussed in Appendix 9.B. In our relative model, however, we set $\lambda^{(R,j)} = 1\ \forall j$, as opposed to $\lambda^{(R,j)} + \lambda^{(S,j)} = 1\ \forall j$ as in Villegas and Haberman (2014). Our identifiability constraint implicitly establishes a hierarchy between the populations, with population S subordinate to population R. Setting $\lambda^{(R,i)} = 1$ motivates the two-stage fitting process, with the age/period and cohort terms being fitted using data for the reference population alone. In our context, as the two populations are of very different sizes, this is both reasonable and unlikely to make a material difference to the fitted parameters. However, it means that the fitted parameters for the sub-population are conditional on those found for the reference model. It is, therefore, important that tests of the model include full allowance for parameter uncertainty in both populations.

As with the model of Russolillo et al. (2011) and the analysis of Villegas and Haberman (2014), it is also possible to apply our model to multiple sub-populations, such as those from different pension schemes. In this case, separate scaling factors would be required for each scheme. For multiple schemes, the hierarchical structure of the model is an advantage, since each scheme can be considered separately once the reference population has been estimated.

9.4 Applying the relative model to SAPS data

9.4.1 The reference models for UK data

Our first task is to construct suitable mortality models for men and women in the national UK population. To do this, we apply the GP to data from the Human Mortality Database (Human Mortality Database (2014)) for the period 1950 to 2011 and for ages 50 to 100. The GP produces a model with three age/period terms, described in Table

9.1,[11] plus cohort terms for both men and women in the UK. All of these terms are shown in Figures 9.1 and 9.2. Further details of the age functions used in this model and tests of the goodness of fit to data are given in Appendix 9.C.

Term	Men		Women	
	Description	Demographic Significance	Description	Demographic Significance
$f^{(R,1)}(x)\kappa_t^{(R,1)}$	Constant age function	General level of mortality	Constant age function	General level of mortality
$f^{(R,2)}(x)\kappa_t^{(R,2)}$	"Call" age function	Older age mortality	"Call" age function	Old age mortality
$f^{(R,3)}(x)\kappa_t^{(R,3)}$	"Put" age function	Younger age mortality	Gaussian age function	Younger age mortality

TABLE 9.1: Terms in the reference models constructed using the general procedure for UK men and women ages 50 to 100

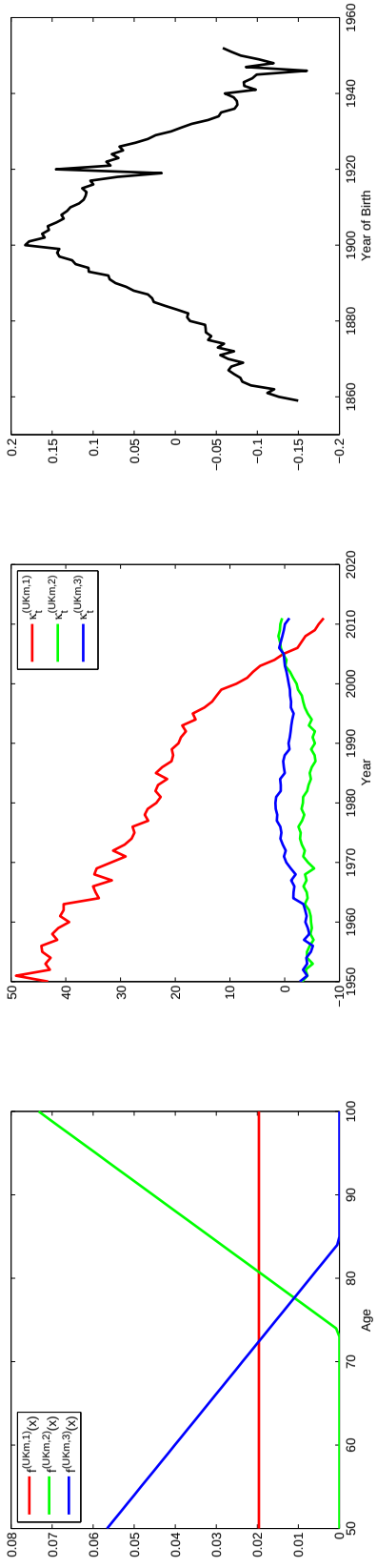
In Figures 9.1c and 9.2c, the most notable features of the cohort parameters for both men and women are the presence of large outliers in 1919/20 and 1946/47. We believe, based on the analysis of Richards (2008), that these are not genuine cohort effects, but are merely data artefacts arising from the surge of births following the large-scale demobilisations after the First and Second World Wars, which biases the calculation of the exposures to risk in the UK population data for those years. We do not expect to find similar outliers in the SAPS data as this is based on aggregating individual scheme-member data rather than population level estimates.[12] One method to solve this would be to adjust the UK population exposures data as proposed in Cairns et al. (2014). However, for simplicity, we choose to retain the original data and employ indicator variables to remove the impact of outliers from the relevant cohort parameters. These adjusted cohort parameters are then used in the analysis which follows.[13]

As discussed in Chapters 3 and 4, many mortality models are not fully identified. To uniquely specify the parameters, we impose identifiability constraints. These constraints

[11] Demographic significance, as used in Table 9.1, is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

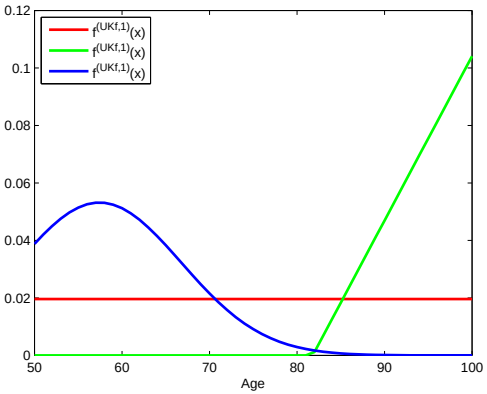
[12] This is borne out by using simple APC models fitted to the SAPS data, which show cohort parameters without these outliers.

[13] It is interesting to note that these outliers may impact the effectiveness of hedging strategies which use securities indexed to national population data, as the index will continue to show a large (but fictitious) effect for specific cohorts which will not be observed in the specific population being hedged. It is therefore important that any indices use national population data which has been adjusted to remove these data artefacts, possibly using the approach of Cairns et al. (2014).

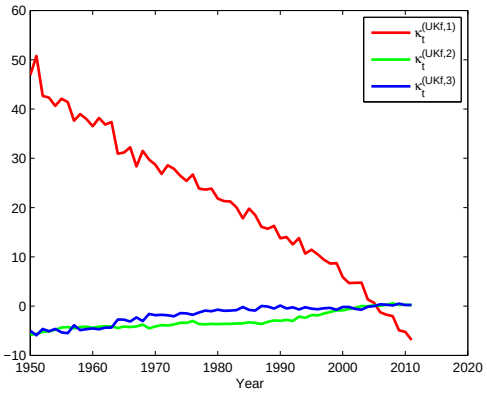


(A) Age functions (B) Period functions (C) Cohort function

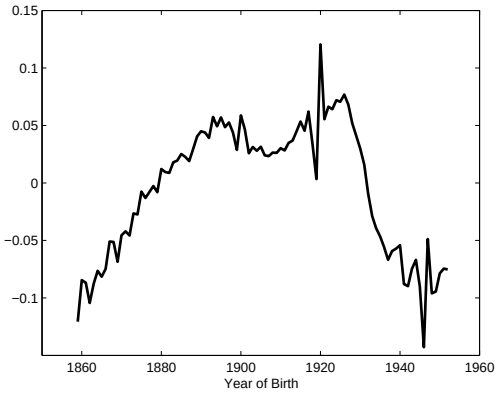
FIGURE 9.1: Age, period and cohort functions in the reference model for men in the UK



(A) Age functions



(B) Period functions



(C) Cohort function

FIGURE 9.2: Age, period and cohort functions in the reference model for women in the UK

are arbitrary, in the sense that they do not affect the fit to data. However, they can be used to impose our desired demographic significance on the parameters.

Models generated by the GP impose the following standard identifiability constraints

$$\sum_{x=50}^{100} |f^{(R,i)}(x)| = 1 \; \forall i, \quad R = \{\text{UKm, UKf}\} \tag{9.6}$$

on the age functions to ensure that they have a consistent normalisation scheme. This enables us to compare the magnitudes of the period functions both with each other and between populations and gauge their relative importance.[14]

In order to assist the visual comparison between the UK and SAPS data (the latter of which only spans ages 60 to 90 and years 2000 to 2011), we impose the following constraint on the period functions

$$\sum_{t=2000}^{2011} \kappa_t^{(R,i)} = 0 \; \forall i, \quad R = \{\text{UKm, UKf}\} \tag{9.7}$$

This means that the period functions represent deviations from an "average" level of mortality in the period covered by the SAPS data, rather than over the whole period of the UK data.

The results of Chapter 4 also indicate that we need to impose constraints on the levels and linear trends present in the cohort parameters. To identify their levels, we impose the following constraints on the cohort parameters for each of the reference populations

$$\sum_{y=1910}^{1951} n_y^{(S)} \gamma_y^{(R)} = 0, \quad R = \{\text{UKm, UKf}\} \tag{9.8}$$
$$S = \{\text{SAPSm, SAPSf}\}$$

where $n_y^{(S)}$ is the number of observations of each cohort in the SAPS data. As with the period functions, this means that the cohort parameters should be centred around zero over the range of the SAPS data, not the full range of the data covered for the UK

[14] For both women and men, the second and third age/period terms use age functions which are "self-normalising" in the sense of Chapter 3.

population. To constrain the linear trends in the cohort parameters, we impose

$$\sum_{y=1850}^{1961} n_y^{(R)} \gamma_y^{(R)} (y - \bar{y}) = 0, \quad R = \{\text{UKm, UKf}\} \tag{9.9}$$

where $n_y^{(R)}$ is the number of observations of each cohort in the UK national data.

The justification for these constraints is that they allow us to remove linear trends in the cohort parameters. This makes them conform better to the demographic significance for cohort parameters described in Chapter 2, namely that the cohort parameters should not have any long-term systematic trends. We impose this over the whole range of the UK data, which is considerably longer than the range covered by the SAPS data, as there appear to be short-term trends (lasting for a few decades) which are then reversed out over a longer time horizon. However, this means that over the shorter range of years of birth covered by the SAPS data, the cohort parameters appear to have strong, negative trends.

It is important to note, however, that our demographic significance for the parameters is highly subjective and our choice of constraints is arbitrary. We have therefore taken appropriate steps in Appendix 9.B to ensure that our choice of identifiability constraints does not affect either the mortality rates fitted by the relative model or our overall conclusions.

9.4.2 The relative models for the SAPS data

We now estimate the relative model using these reference age, period and cohort terms for the full SAPS dataset. As discussed in Section 9.3, we do this in stages using a specific-to-general procedure. We start with the simplest and most restricted model, i.e., where $\alpha_x^{(\Delta)}$ is restricted to take a parametric form and we restrict the scaling factors $\lambda^{(j)}$ to be equal to zero. This model is referred to as Model 1 in Tables 9.2 and 9.3 below.

We then allow these restrictions to be relaxed sequentially. This means that, in turn, we estimate the relative model with all possible combinations of constraints, where $\alpha_x^{(\Delta)}$ is either parametric or non-parametric and where $\lambda^{(j)}$ can be restricted to be equal to zero, unity or allowed to vary freely. This gives us $162 (= 2 \times 3^4)$ different combinations of constraints for the two alternative structures for $\alpha_x^{(\Delta)}$ and three alternatives for each of the four different scaling factors, $\lambda^{(j)}$. For each of these different models, the goodness

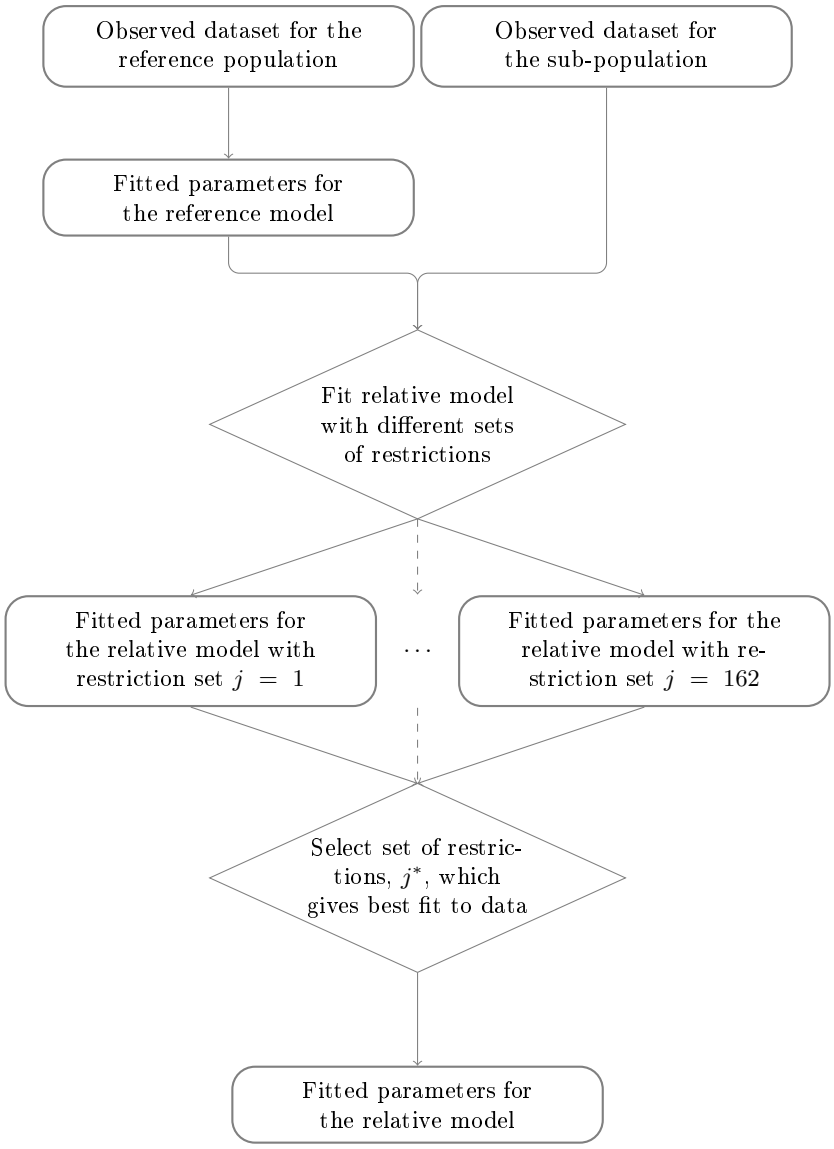


FIGURE 9.3: Flow chart illustrating the procedure for fitting and selecting the relative model

of fit to the data is calculated, as measured by the BIC. The model which gives the best fit to data (i.e., the highest BIC) is then selected as the preferred model, referred to as Model 8 in Tables 9.2 and 9.3, for the dataset. This process is illustrated in Figure 9.3.

Several of the models tested, with representative combinations of restrictions, are shown in Tables 9.2 and 9.3 for the male and female SAPS data.[15] These have been chosen to illustrate the impact of relaxing various restrictions, for instance, comparing Models 1 and 2 illustrates the impact on the goodness of fit of using a non-parametric as opposed to a parametric structure for $\alpha_x^{(\Delta)}$, whilst comparing Models 3 and 4 illustrates the impact of introducing the set of cohort parameters from the reference population.

[15] In Tables 9.2 and 9.3, "NP" stands for non-parametric while "P" stands for parametric.

The preferred model which maximises the fit to data is shown as Model 8. However, it is important to note that the fitting procedure tests all 162 possible combinations for the structure of $\alpha_x^{(\Delta)}$ and any combination of restrictions on $\lambda^{(j)}$.

Model No.	1	2	3	4	5	6	7	8
$\alpha^{(\Delta)}$	P	NP	P	P	NP	P	NP	P
$\lambda^{(1)}$	0	0	1	1	1	1.36	1.37	1.35
$\lambda^{(2)}$	0	0	1	1	1	1.78	1.93	1.73
$\lambda^{(3)}$	0	0	1	1	1	2.01	1.97	2.00
$\lambda^{(\gamma)}$	0	0	0	1	1	0.86	0.51	1
Log-likelihood $\times 10^3$	-2.14	-2.06	-2.00	-1.94	-1.89	-1.91	-1.86	-1.92
Free parameters	3	31	3	3	31	7	35	6
BIC $\times 10^3$	-2.15	-2.15	-2.01	-1.95	-1.98	-1.93	-1.96	-1.93

TABLE 9.2: Representative sets of restrictions for the relative model using male SAPS data

Model No.	1	2	3	4	5	6	7	8
$\alpha^{(\Delta)}$	P	NP	P	P	NP	P	NP	P
$\lambda^{(1)}$	0	0	1	1	1	1.24	1.20	1.22
$\lambda^{(2)}$	0	0	1	1	1	2.35	2.45	2.42
$\lambda^{(3)}$	0	0	1	1	1	0.09	-0.06	0
$\lambda^{(\gamma)}$	0	0	0	1	1	1.06	0.97	1
Log-likelihood $\times 10^3$	-2.05	-2.01	-1.94	-1.83	-1.80	-1.80	-1.77	-1.80
Free parameters	3	31	3	3	31	7	35	5
BIC $\times 10^3$	-2.06	-2.10	-1.95	-1.83	-1.89	-1.82	-1.87	-1.82

TABLE 9.3: Representative sets of restrictions for the relative model using female SAPS data

For both men and women, the preferred model selects a parametric simplification for the difference in the level of mortality, $\alpha_x^{(\Delta)}$. This substantially reduces the number of free parameters in the preferred model, leading to greater parsimony. This is also borne out by comparing models which differ by the form of $\alpha_x^{(\Delta)}$, but have similar restrictions placed on the scaling factors, $\lambda^{(j)}$, e.g., Models 1 and 2, or Models 4 and 5 in Tables 9.2 and 9.3. In some respects, this supports the traditional actuarial practice of adjusting mortality rates for a pension scheme by taking a mortality table from a reference population (in this case, the full UK population) and making relatively simple adjustments to it. We also see from Figures 9.5a and 9.5b that $\alpha_x^{(\Delta)}$ is generally negative across all ages. This indicates that the SAPS population has generally lower levels of mortality rates than the national population, which is consistent with the results of Continuous Mortality Investigation (2011).

In the case of the male data, the procedure selects a model where all the $\lambda^{(i)}$ for the age/period terms are allowed to vary freely, i.e., without any restrictions placed upon them at the estimation stage. The same is true for the female data, except that $\lambda^{(3)}$ is set to be equal zero. This is unsurprising given the other models shown in Table 9.3: in the models where $\lambda^{(3)}$ for women is allowed to vary (e.g., Models 6 and 7), it takes a value comparatively close to zero, and so it can be restricted to equal zero without adversely affecting the goodness of fit of the model.

We also see that the scaling factors for the period functions for both men and women are greater than unity when their estimation is not restricted. This indicates that the SAPS populations are responding to the same drivers of mortality rates as the national population, but with greater sensitivity to these underlying causes. Since mortality rates are generally falling in the UK, this implies that the rate of improvement in longevity is slightly faster for members of occupational pension scheme than for the national population. This contrasts with the findings of Continuous Mortality Investigation (2011), which found that the falls in standardised mortality ratios for the SAPS populations broadly mirrored the falls observed in the wider UK population. However, since the standardised mortality ratio is an aggregate measure of mortality, which takes account of the level of mortality rates, it is likely that the difference between our results and those of Continuous Mortality Investigation (2011) are not significant.

In addition, for both sexes, $\lambda^{(\gamma)}$ is restricted to be equal to unity. This means that we do not expect any systematically different cohort effects in the SAPS data compared to those observed in the reference population. It is interesting to compare this to the results of Li et al. (2013), which also found that the cohort effects in related populations can often be assumed to be equal to each other without adversely affecting the goodness of fit for a model.

Finally, we note that the BICs of many of the models with different restrictions are very similar, meaning that there is not much to choose between them. It may therefore be justifiable to select simpler models than suggested by looking just at goodness of fit, on the grounds that they may be more robust to parameter uncertainty or easier to project into the future, as done in Section 9.5. This will be even more important when we investigate smaller, pension scheme-sized datasets, as in Section 9.6.

9.4.3 Parameter uncertainty and model risk

We next consider the robustness of the preferred model selected, i.e., Model 8. We do this in two stages, by considering the different sources of uncertainty outlined in Cairns (2000). First, we consider only parameter uncertainty, i.e., the uncertainty in the free parameters of the preferred model, on the assumption that the restrictions placed on the parameters in Model 8 are correctly specified. Second, we allow for model risk by allowing the procedure to select different models using the sequential procedure discussed above.

For both stages, we use a procedure based on the residual bootstrapping method of Koissi et al. (2006) to generate new pseudo-data. This resamples from the fitted residuals to generate new simulated death counts to which the model is refitted, allowing the uncertainty in the parameters to be measured. We do this first to allow for parameter uncertainty in the reference model. It is important to allow for parameter uncertainty in the reference model due to the hierarchical structure of the relative model, i.e., that the parameters for the reference model are implicitly assumed to be known when the relative model is fitted. Therefore, uncertainty in the parameters of the reference model can be magnified when we come to investigate the uncertainty in the parameters of the relative model.

The next step is to bootstrap new pseudo-data for the sub-population. When using a residual bootstrapping procedure, it is important that the fitted residuals being used contain as little structure as possible, so that very little of the information in the original data is lost when these residuals are randomly resampled. This will be the case for models which provide a close fit to the data (i.e., a high maximum likelihood), irrespective of the number of free parameters used by the model to achieve this fit. Therefore, in our residual bootstrapping procedure we use the expected mortality rates and fitted residuals from Model 7, since this model has the highest log-likelihood in Tables 9.2 and 9.3 above. However, since Model 7 is outperformed by a number of other models when the goodness of fit is penalised for the number of parameters, we do not specifically consider it further.

9.4.3.1 Parameter uncertainty

For the first stage, we consider only parameter uncertainty. To do this, we fit the relative model to 1,000 sets of pseudo death counts, generated by the Koissi et al. (2006)

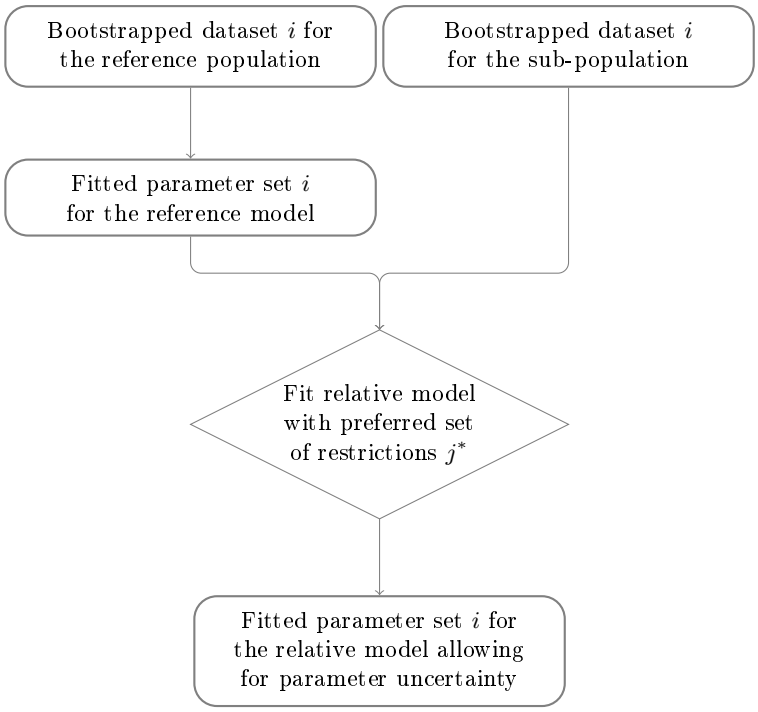
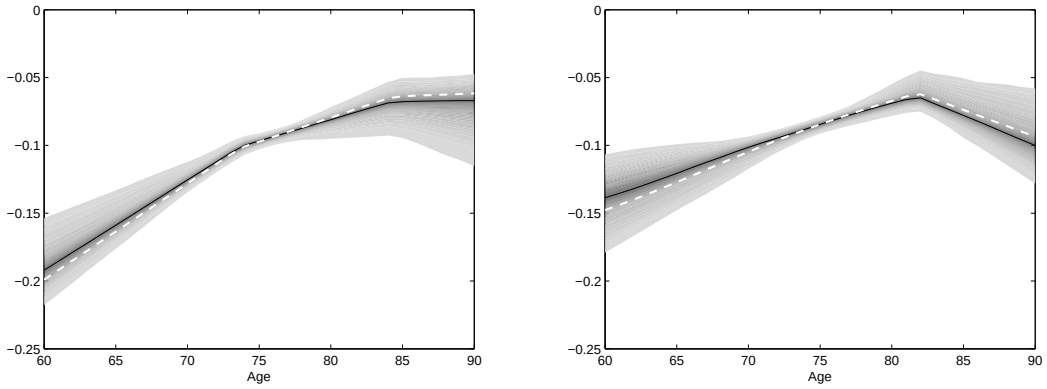


FIGURE 9.4: Flow chart illustrating the procedure for fitting and selecting the relative model allowing for parameter uncertainty



(A) Male (B) Female

FIGURE 9.5: 95% fan charts showing the level of parameter uncertainty in $\alpha_x^{(\Delta)}$

residual bootstrapping procedure. For each of these datasets, however, we do not test which set of restrictions give the best fit to the data. Instead we impose the same set of restrictions as were used for Model 8 in Tables 9.2 and 9.3. We fit the relative model with the restrictions in Model 6 (which allows all scaling factors to freely vary) used as a comparator. This process is illustrated in Figure 9.4.

Figure 9.5 shows the impact of parameter uncertainty on the level parameters by showing the 95% fan chart. To interpret this, we note that a 95% confidence interval for $\alpha_x^{(\Delta)}$

of width 0.1 at age, x, (i.e., $\alpha_x^{(\Delta)} \in (\hat{\alpha}_x^{\Delta} - 0.1, \hat{\alpha}_x^{\Delta} + 0.1)$) roughly corresponds to a 95% confidence interval for the fitted mortality rates of $(0.90\hat{\mu}_{x,t}, 1.10\hat{\mu}_{x,t})$, where $\hat{\mu}_{x,t}$ is our best estimate of the mortality rate. For comparison, differences in the level of mortality of around this order of magnitude are visible between different industrial sectors in Figures 8 and 9 of Continuous Mortality Investigation (2012). This implies that it may be difficult to robustly determine differences in the level of mortality between individuals who worked in different industries once parameter uncertainty is taken into account. The dashed lines in Figure 9.5 show the parameter-certain estimates of $\alpha_x^{(\Delta)}$, which lie close to the centre of the confidence intervals given by relative models.[16]

	Men		Women	
	Model 6	Model 8	Model 6	Model 8
$\lambda^{(1)}$	[1.21,1.40]	[1.23,1.39]	[1.13,1.31]	[1.12,1.31]
$\lambda^{(2)}$	[1.47,1.93]	[1.44,1.91]	[1.66,2.96]	[1.79,2.94]
$\lambda^{(3)}$	[1.44,2.33]	[1.45,2.33]	[-0.56,0.82]	0
$\lambda^{(\gamma)}$	[0.71,1.09]	1	[0.87,1.22]	1

TABLE 9.4: 95% confidence intervals for scaling factors in Model 6 and Model 8 fitted to male and female SAPS data

Table 9.4 shows the 95% confidence intervals for the scaling factors for men and women. The first thing to note from these results is that the scaling factors are subject to substantial parameter uncertainty. As the relative model is very parsimonious and contains relatively few free parameters, this should caution us against considering more sophisticated models for the SAPS population. For instance, we are unlikely to have sufficient data to robustly estimate separate period functions for the SAPS data compared with the reference population, which was done in Villegas and Haberman (2014).

We also note that the confidence intervals for $\lambda^{(1)}$ tend to be slightly narrower than those for the other age/period terms, which is to be expected since the first age function covers the entire age range and therefore the estimate of $\lambda^{(1)}$ uses more data. This cautions us against trying to estimate age-specific or year-specific parameters, since these would have relatively little data to support them, leading to substantial parameter uncertainty in their estimates. For instance, the uncertainty in the estimate of a non-parametric form for $\alpha_x^{(\Delta)}$ would be considerably higher, since this requires separate parameters at each age to be estimated.

[16] This indicates that our method for estimating parameter uncertainty does not significantly bias the results, which is an important check of its suitability.

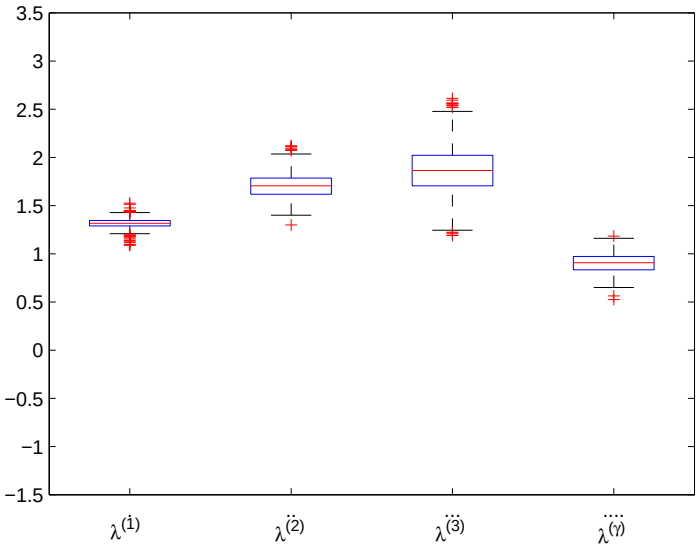
From Table 9.4, we can easily apply a simple but important check of our modelling approach by using an alternative method for determining suitable restrictions of the relative model such as a "general-to-specific" approach described in Campos et al. (2005). This would fit an unrestricted model (i.e., Model 6) to the data, observe the confidence intervals for each parameter and use these to determine which restrictions to apply. To illustrate, if the confidence interval for $\lambda^{(j)}$ included unity, the general-to-specific approach would impose $\lambda^{(j)} = 1$ on the grounds of statistical significance. From Table 9.4, we see that the confidence intervals for $\lambda^{(\gamma)}$ for both men and women contains zero, whilst the confidence interval for $\lambda^{(3)}$ for women contains unity. Therefore, the general-to-specific approach would arrive at the same set of restrictions for the preferred model as our approach, which is based solely on considering the goodness of fit of the relative model with different sets of restrictions.

We also see, by comparing the confidence intervals for the unrestricted parameters in Model 8 with their counterparts from Model 6, that imposing the preferred set of restrictions does not significantly affect the estimation of the other parameters in the model. This, again, acts as a useful check to ensure that the procedure we have used to select the preferred set of restrictions does not remove statistically significant parameters from the relative model, and gives us confidence that our approach merely removes unnecessary parameters and so leads to a more parsimonious model.

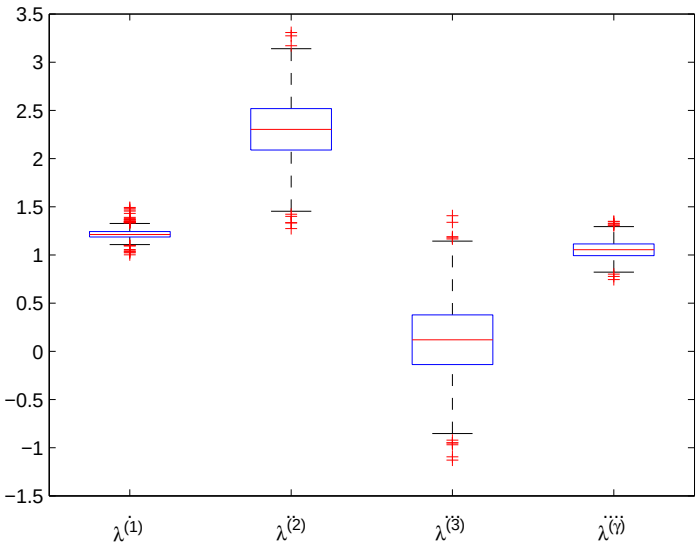
Inspection of the boxplots of the bootstrapped parameters from Model 6, shown in Figure 9.6, indicates that the confidence intervals appear roughly symmetric around their midpoints. However, on closer inspection, $\lambda^{(1)}$ shows substantial skewness. Investigating this further, Jarque-Bera tests on the bootstrapped rejects the assumption of normality for $\lambda^{(1)}$ for both sexes and for $\lambda^{(3)}$ for women at the 5% level. This indicates that we cannot reliably use asymptotic methods based on the information matrix (similar to those used in Brouhns et al. (2002b)) to allow for parameter uncertainty, since these methods assume that the parameters will be normally distributed. This justifies the use of residual bootstrapping procedures, such as the one proposed here, in order to properly investigate parameter uncertainty in these models.

9.4.3.2 Model risk

The second stage of testing the robustness of the model is to fit the relative model to the bootstrapped data without specifying the form of the preferred model. Instead, we allow the procedure to select a potentially different preferred model in each simulation.



(A) Men



(B) Women

FIGURE 9.6: Boxplots of the bootstrapped parameters from Model 6

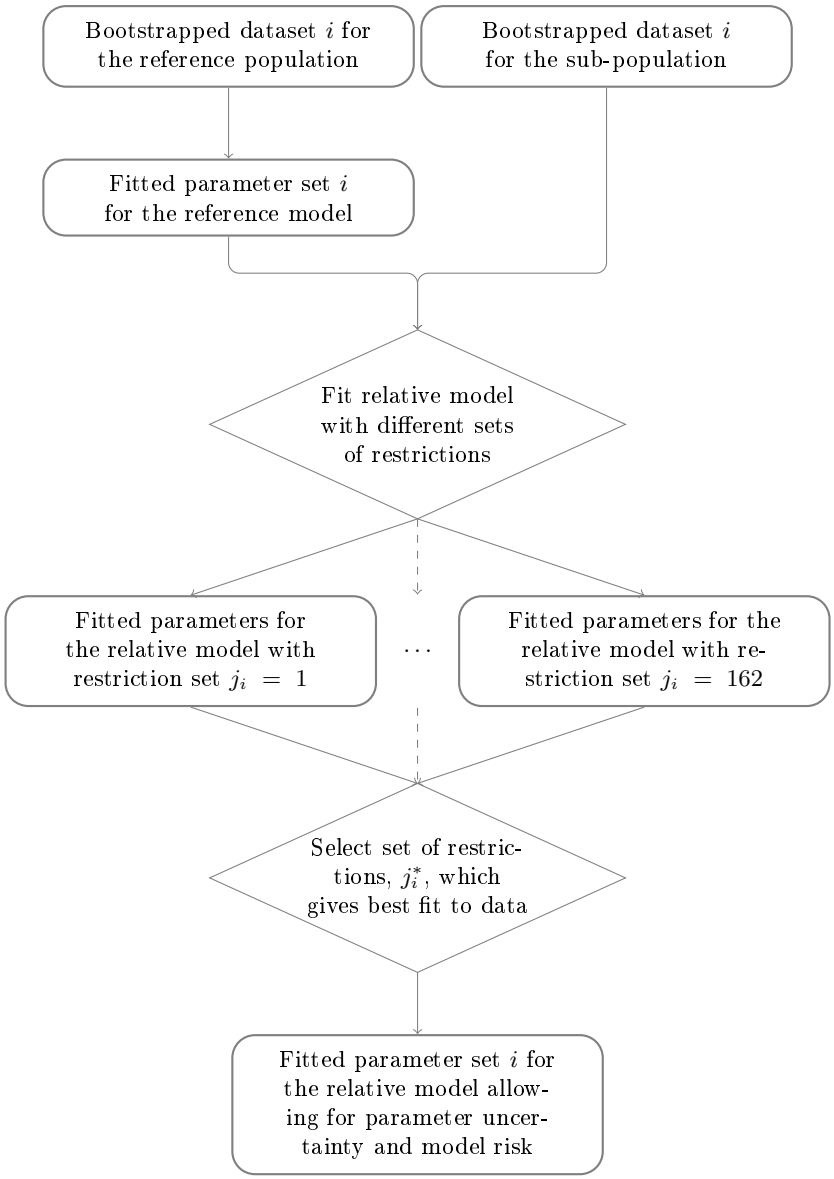


FIGURE 9.7: Flow chart illustrating the procedure for fitting and selecting the relative model allowing for parameter uncertainty and model risk

This allows for "model risk", in the sense of Cairns (2000), i.e., the risk that the model selected is not an accurate representation of the true processes generating the data. This process is illustrated in Figure 9.7. However, we are still selecting a preferred model from a relatively limited set of comparators, and so the procedure does not fully capture the potential for model risk.

Looking first at the preferred form of $\alpha_x^{(\Delta)}$, we find that, from 1,000 bootstrapped datasets, we find that the preferred model restricts $\alpha_x^{(\Delta)}$ to have a parametric form in 88% of the datasets for men and 100% of the datasets for women. The modelling approach, therefore, overwhelmingly prefers imposing a parametric structure for $\alpha_x^{(\Delta)}$ over

allowing this to vary freely, even when allowing for model risk. This is not too surprising when considering the results in Tables 9.2 and 9.3, as these showed that allowing $\alpha_x^{(\Delta)}$ to take a non-parametric form significantly worsened the goodness of fit when this was penalised for the additional number of parameters.

Restriction placed on:		$\lambda^{(j)} = 0$	$\lambda^{(j)} = 1$	$\lambda^{(j)}$ unrestricted
Men:	$\lambda^{(1)}$	0%	67%	33%
	$\lambda^{(2)}$	0%	0%	100%
	$\lambda^{(3)}$	0%	36%	64%
	$\lambda^{(\gamma)}$	1%	71%	28%
Women:	$\lambda^{(1)}$	0%	69%	31%
	$\lambda^{(2)}$	0%	2%	98%
	$\lambda^{(3)}$	93%	7%	0%
	$\lambda^{(\gamma)}$	0%	97%	3%

TABLE 9.5: Frequency of different restrictions being placed upon the scaling factors in the preferred relative model, based on 1,000 bootstrapped datasets

Table 9.5 shows the frequency of observing the various restrictions on the scaling factors in the preferred model based on the same 1,000 bootstrapped datasets. We note that the most likely form that these restrictions take is the preferred one found for Model 8 in Tables 9.2 and 9.3. The exception to this is $\lambda^{(1)}$ for both men and women: when model risk is allowed for, the most likely outcome is that $\lambda^{(1)}$ is restricted to equal unity, while this parameter was allowed to vary freely in Model 8 for both sexes. We are unsure why this should be the case. However, we note that it is inevitable that some information in the original data will be lost due to the random resampling of the fitted residuals in the Koissi et al. (2006) approach. Therefore, it is likely that the preferred model for bootstrapped data will be simpler and have more restrictions placed upon it, as fewer parameters will be required to capture the reduced level of information in the bootstrapped data compared with the original data.

In summary, we find that there is substantial model risk for both sexes, and no one set of restrictions out of the available options is universally selected. This will be important when we project the model in Section 9.5. It should also, again, caution us against using overly complicated models for the SAPS populations, as there is substantial uncertainty not only in any parameter estimates found but also in the fundamental form of the model.

9.5 Basis risk and projecting mortality for the SAPS population

In Section 9.4, the relative model was applied to historical data for the SAPS population. Given projections of the reference population, we can also use the relative model to map these into projections for the sub-population.

Many pension schemes are concerned about "basis risk", the risk that the mortality experience of the scheme in question will be substantially different to that of the national population. This is important when assessing hedging strategies (for instance, in Li and Hardy (2011), Coughlan et al. (2011) and Cairns et al. (2013)) using financial instruments based on national mortality rates. More fundamentally, it is an important question when funding a pension scheme, since most standard projections for future mortality rates are based on analysing national populations (for instance, the CMI mortality projection model in Continuous Mortality Investigation (2009a) which is widely used in the UK).

Intuitively, basis risk can arise because of a difference in levels of mortality rates (e.g., the specific population exhibiting systematically higher or lower mortality rates than the reference population as a result of characteristics such as socio-economic status which will change only slowly) and a difference in trends in mortality rates (i.e., mortality rates evolving differently in the sub-population, for instance, due to preferential access to new medications) between the two populations. In terms of the relative model of Equation 9.3, these can be thought of as relating to $\alpha_x^{(\Delta)}$ and the $\lambda^{(j)}$, respectively. Level differences can be measured relatively easily using traditional actuarial methods which are well within the capabilities of modern scheme actuaries. However, the difference in trends between populations is more difficult to measure reliably and, consequently, is of greater concern to many scheme actuaries.

In order to evaluate the potential impact of basis risk between the UK and SAPS populations, we first need to project mortality rates for the national population. However, it is important that our projections of mortality rates are "well-identified" in the sense of Chapters 3 and 4 in that they do not depend upon our chosen identifiability constraints. To project the reference population, we therefore adopt the techniques of Chapter 4 and use random walks with drift

$$\boldsymbol{\kappa}_t^{(R)} = \boldsymbol{\mu}^{(R)} + \boldsymbol{\kappa}_{t-1}^{(R)} + \boldsymbol{\epsilon}_t^{(R)} \tag{9.10}$$

where $\boldsymbol{\kappa}_t^{(R)} = \left(\kappa_t^{(R,1)}, \quad \dots \quad \kappa_t^{(R,N)}\right)^\top$, $\boldsymbol{\mu}^{(R)}$ are drift coefficients and $\boldsymbol{\epsilon}_t^{(R)}$ are normally distributed, contemporaneously correlated innovations. For the cohort parameters, we make projections using an AR(1) around "well-identified" drifts

$$\gamma_y^{(R)} - \beta^{(R)} X_y = \rho^{(R)} (\gamma_{y-1}^{(R)} - \beta^{(R)} X_{y-1}) + \varepsilon_y \tag{9.11}$$

where X_y is a vector of deterministic functions[17] and $\beta^{(R)}$ are drift coefficients.

The deterministic functions, X_y, are chosen to ensure that the projections are "well-identified", i.e., that the projected mortality rates for the reference population do not depend upon the identifiability constraints used when fitting the model. To achieve this in the context of the reference models developed in Section 9.4.1 and Appendix 9.C, we have

$$\begin{aligned} X_y &= \left(1, \quad (y - \bar{y})\right)^\top \\ \beta^{(R)} &= \left(\beta_0^{(R)}, \quad \beta_1^{(R)}\right) \quad R = \{\text{UKm, UKf}\} \end{aligned}$$

Any dependence between mortality rates for men and women is not relevant to the following discussion, where only the relationships between mortality rates in the reference and sub-populations for the same sex are investigated. Therefore, in these projections, we do not take into account any dependence between male and female mortality rates in the reference population, and consequently project these populations independently. A more complete analysis of the mortality and longevity risks in pension schemes, such as in Chapter 10 , would need to allow for dependence between sexes in the reference population. For techniques which could allow for dependence between these populations, see Chapter 8 and the references therein.

To illustrate the basis risk between the SAPS and UK populations, we consider annuity values at age 65 (calculated using a real discount rate of 1% p.a.). We perform 1,000 Monte Carlo simulations using the time series processes above to give projected mortality rates in the national population, which are then used to generate projected mortality rates in the SAPS population using the relative mortality models for men and women separately. Basis risk is accounted for using the relative model in three stages:

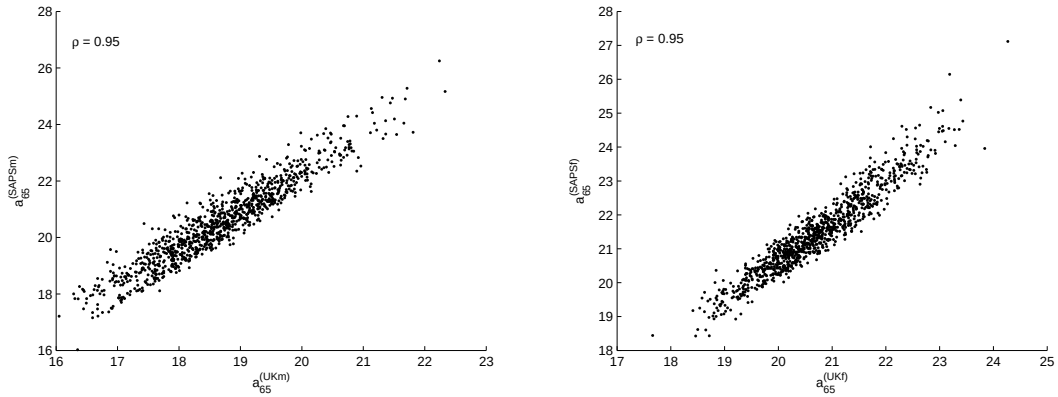
[17] We have used the same notation for the trends, X_y, in Equation 9.11 as was used for the additional functions of year of birth in the relative model in Equation 9.3. However, the reader should be aware of the slight difference in definition between these two contexts, namely that in Equation 9.11, $X_y = \left(1, \quad (y - \bar{y})\right)^\top$, whilst in Equation 9.3, $X_y = \left(y - \bar{y}\right)^\top$, i.e., X_y did not possess a constant.

1. First, we allow only for the impact of the random innovations, $\boldsymbol{\epsilon}_t^{(R)}$ and $\varepsilon_y^{(R)}$, on projected mortality rates, i.e., we allow for process risk in the terminology of Cairns (2000). We do this by using Equations 9.10 and 9.11 to project stochastically the period and cohort parameters found for the reference population in Section 9.4.1, and then using the preferred relative model estimated in Section 9.4.2 and shown as Model 8 in Tables 9.2 and 9.3. Using this technique, we find correlations between annuity values in the UK and SAPS populations of 99% for men and 98% for women.

2. Second, we allow for parameter uncertainty in both populations. To do this, we use the approach illustrated in Figure 9.4 to generate new parameters for both the reference and the sub-populations. The time series processes in Equations 9.10 and 9.11 are then re-estimated for the bootstrapped period and cohort parameters for the reference model, and mortality rates for the reference and sub-populations projected from these. When allowing for parameter uncertainty, we find correlations between annuity values in the UK and SAPS populations of 98% and 97% for men and women, respectively, indicating than parameter uncertainty has not added significantly to the basis risk between the two populations. This is surprising, given the results of Section 9.4.3.1 as shown in Figures 9.5 and 9.6, which showed relatively high levels of uncertainty in the levels and scaling parameters. However, this may indicate that the basis risk arising from different rates of change in mortality in different populations may not be particularly significant, as discussed in Section 9.7.

3. Finally, we allow for model risk in the selection of the preferred model for the sub-population. We do this using the same procedure as illustrated in Figure 9.7 to generate new parameters for the reference population and a new preferred model for the sub-population. The time series processes in Equations 9.10 and 9.11 are then re-estimated for the bootstrapped period and cohort parameters for the reference model, and mortality rates for the reference and sub-populations projected from these. Using this procedure, we observe correlations between annuity values in the UK and SAPS populations of 95% for men and 96% for women. It is interesting to note that for both sexes, we achieve correlations of over 90%, even when allowing for all three sources of uncertainty in the relative model.

Note that this analysis looks only at annuity values (i.e., the expected present value of payments to an individual) and so does not consider the idiosyncratic risk that would also be present in the benefits payable from a pension scheme. This was investigated in Donnelly (2014), Aro (2014) and, in particular, in Chapter 10 where we find this is likely



(A) Men (B) Women

FIGURE 9.8: Projected annuity values for the UK and SAPS populations from 1,000 Monte Carlo simulations

to be substantial for even relatively large pension schemes.

Figure 9.8 shows scatter plots of annuity values calculated using mortality rates in the UK and SAPS populations for men and women in the third, most general case (i.e., incorporating process risk, parameter uncertainty and model risk). It is interesting to note that, for both sexes, the systematic longevity risk (indicated by the range of values the annuity value can take, e.g., 18 to 24 in the case of men) is far greater than the basis risk. Indeed, the systematic longevity risk accounts for around 90% of the uncertainty in an annuity value for the SAPS population, indicating that basis risk may be considerably less important than is commonly believed. This is discussed further in Section 9.7.

However, it is important to note that in all of these cases, there is no genuine trend basis risk between the two populations. This is because the same processes, i.e., $\boldsymbol{\kappa}_t^{(R)}$ and $\gamma_y^{(R)}$, control the evolution of mortality in both populations, albeit scaled by uncertain factors in the sub-population. This helps explain why the correlations we find are somewhat higher than those found in other studies of basis risk, such as Cairns et al. (2013). However, we note that most of these studies used sub-populations which were considerably larger and covered a longer period of time than the SAPS population. Consequently, there is a trade-off. On the one hand, we might wish to use more complicated models that might give a more accurate assessment of basis risk, but which require larger volumes of data to estimate robustly and, therefore, might involve using data for a larger sub-population which is less relevant for the mortality experience of a specific pension scheme (for instance, the CMI Assured Lives dataset). On the other hand, we might prefer to use simpler models, which can be robustly estimated from smaller datasets that are likely to be more relevant to the specific scheme experience, but give a less accurate

assessment of basis risk. The impact of this trade-off is discussed in Section 9.7.

Finally, the importance of model risk and parameter uncertainty will tend to increase if we consider populations smaller than the SAPS population, as we do in Section 9.6. We would therefore expect to see correlations of a similar size to those found in other studies for population sizes that are more typical of UK pension schemes, due to the greater parameter uncertainty and model risk, even without allowing for genuine trend basis risk. In addition, the cashflows experienced by a pension scheme will also have (potentially substantial) idiosyncratic risk due to the relatively low number of lives under observation. This suggests that, in practice, it would be impossible to distinguish trend basis risk from parameter and model uncertainty for most pension-scheme sized populations. Therefore, any concern about trend basis risk may be misplaced, since it would be impossible to reliably quantify and be small relative to the impact of the other risks in the model. This is discussed further in Section 9.7 and Chapter 10 .

9.6 Applying the relative model to small populations

While the SAPS population is small relative to the national UK population, it does have annual exposures to risk of over one million lives each for men and women, and so still represents a population larger than almost all occupational pension schemes (with the exception of some state schemes). However, the methods developed in this study can be applied to significantly smaller populations, such as those more comparable with the size of large occupational pension schemes.

As discussed in Section 9.4.2, the relative model applied to the SAPS population exhibited a strong preference for parsimony. However, parameter uncertainty and model risk were still important considerations, even with a relatively simple model and the full SAPS data. It is therefore exceedingly likely that in even smaller populations, these considerations will dominate what we can and cannot realistically say about the evolution of mortality of a small sub-population such as that associated with an individual pension scheme.

We investigate the effect of population size on the ability of the relative model to measure mortality differences with the national population by randomly generating scheme-sized exposures to risk and death counts (denoted by lower-case s) based on the SAPS data.

We adopt the following procedure to generate pseudo-data for a scheme with N lives (considering each sex separately):

1. We first rescale SAPS exposures, $E_{x,t}^{(S)}$, to give a proxy for smaller pension schemes with approximately N members. We could, in principle, do this very simply by setting

$$E_{x,t}^{(s)} = E_{x,t}^{(S)} \times \frac{N}{\sum_{\xi} E_{\xi,t}^{(S)}}$$

   This would give a scheme with a constant exposure to risk over each year, but the same pattern of exposures to risk across different ages. However, this simple approach does not capture the pattern of exposures across years seen in the actual SAPS data, due to the partial submission of scheme data in the first and last few years of the SAPS datasets (discussed in Section 9.2, see also Figure 9.13a). This means that, were we to artificially generate a scheme of the same size as the SAPS population, we would not recover the observed SAPS exposures and so would obtain inconsistent results. Since we will apply this procedure to generate pseudo-scheme data for schemes of widely varying sizes, up to and in excess of the full SAPS data, it is essential that our results are consistent with the results we found in previous sections. Consequently, we amend the scaling factors so that

$$E_{x,t}^{(s)} = E_{x,t}^{(S)} \times \frac{5N}{\sum_{\xi} \sum_{\tau=2004}^{2008} E_{\xi,\tau}^{(S)}}$$

   This modifies the denominator to reflect the average exposure to risk in the SAPS data in years 2004-2008, for which almost all relevant pension schemes have submitted data to the SAPS study. This approach therefore replicates the full SAPS data when we generate a scheme of the same size as the SAPS population (including the pattern of relatively low exposures to risk for the first and last years, along with the pattern of exposures at different ages found in the SAPS data.

2. We then generate random death counts for the scheme by modelling them as Poisson random variables. To do this, we use the exposures to risk generated using both the procedure above and the crude mortality rates observed in the SAPS dataset,

$$D_{x,t}^{(s)} \sim Po\left(\frac{D_{x,t}^{(S)}}{E_{x,t}^{(S)}} E_{x,t}^{(s)}\right)$$

We then fit the relative model to this pseudo-scheme data, testing all 162 sets of possible restrictions on the parameters to determine the preferred model using the same procedure described in Section 9.4.3.2. This procedure is illustrated in Figure 9.9. Such

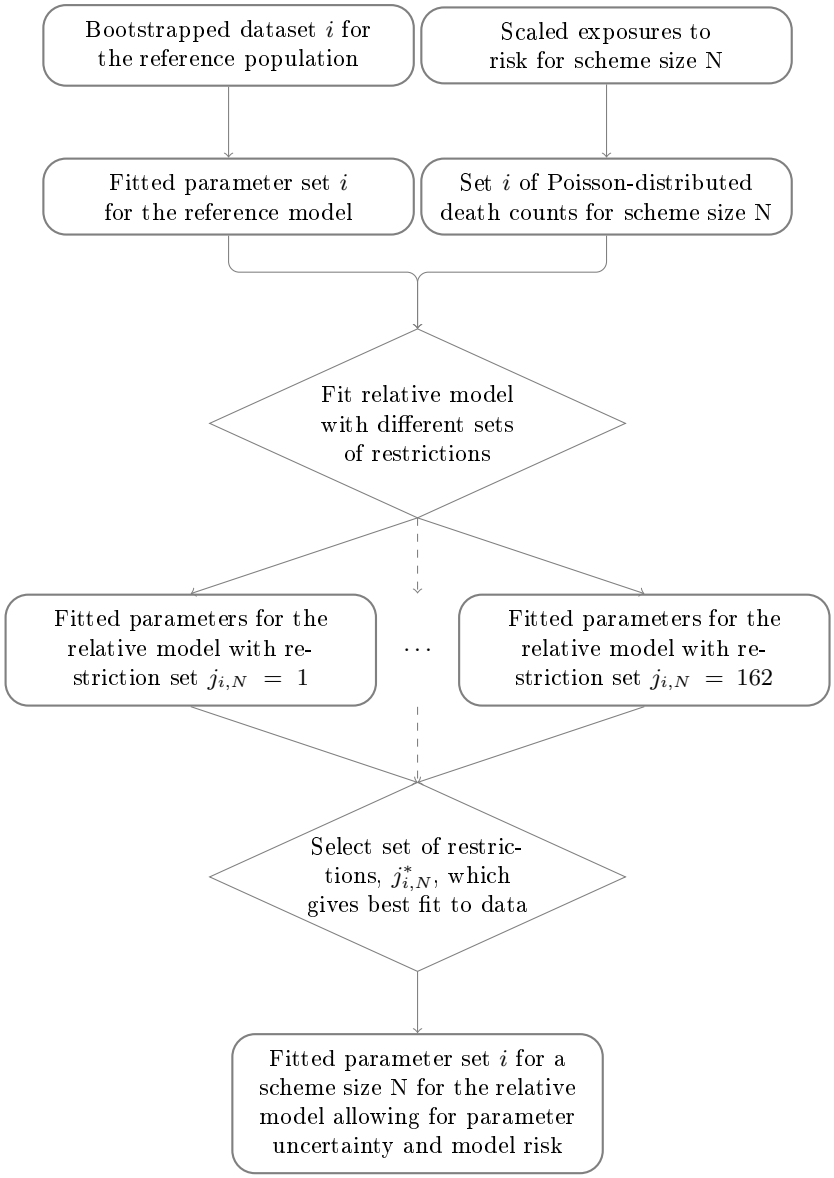


FIGURE 9.9: Flow chart illustrating the procedure for generating data and fitting the relative model to scheme-sized populations, allowing for parameter uncertainty and model risk

an approach is conceptually similar to the "semi-parametric" bootstrapping technique in Brouhns et al. (2005), except we rescale the exposures in order to simulate the range of different scheme sizes present in the UK.

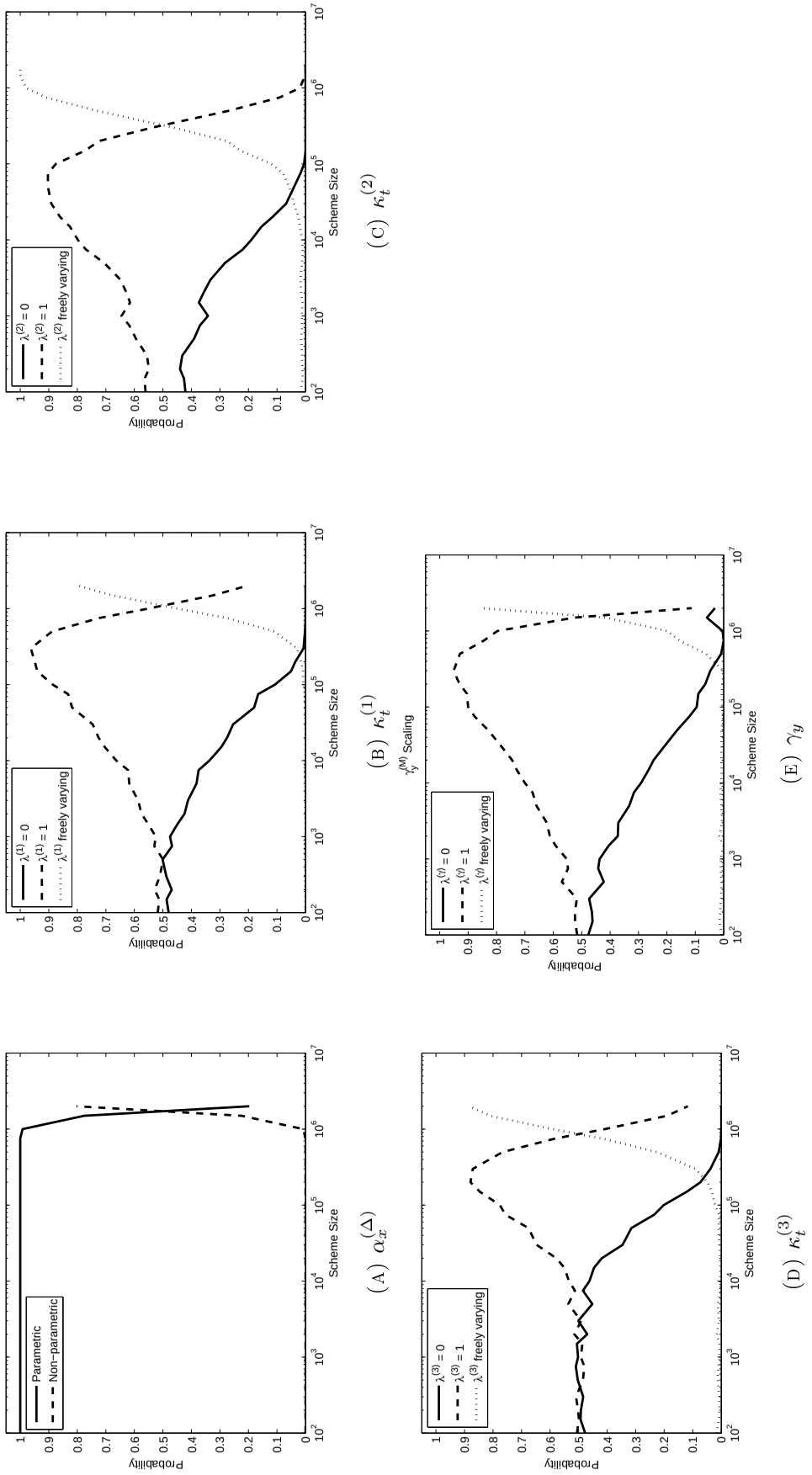
To gain a better understanding of the impact of the size of the population on the complexity of the preferred model, we apply this procedure for scheme sizes at regular intervals in the range $N \in (10^2, 10^6)$ and for 1,000 sets of random death counts at each scheme size. This range of population sizes covers almost the entire range of pension scheme sizes in the UK, and the fitting of multiple models allows for potential model risk in the

selection of the preferred model. The results of this procedure for men and women are shown in Figures 9.10 and 9.11.

First, let us consider the results shown in Figures 9.10a and 9.11a. These figures show that the probability of the procedure preferring a parametric restriction for $\alpha^{(\Delta)}$ is almost unity for schemes up with up to one million members of each sex, which is far in excess of all but the largest state schemes in the UK. This indicates an overwhelming preference for parametric restrictions for $\alpha_x^{(\Delta)}$ in all but the very largest scheme sizes. The implication of this is that making simple adjustments to a standard mortality table will be sufficient to capture the difference in levels in mortality for almost all UK schemes, with little or no need to graduate a bespoke table (even if the data is available).

Looking at the scaling factors for the age/period and cohort terms, we see that, typically, the smallest schemes (fewer than 1,000 members of each sex) are indifferent between restricting $\lambda^{(j)}$ to be equal to zero or unity. For instance, Figure 9.10b shows that the procedure imposes the restriction $\lambda^{(1)} = 0$ and $\lambda^{(1)} = 1$ for men in approximately 50% of the simulations for small schemes, with $\lambda^{(1)}$ being estimated without restrictions in almost no cases. This pattern is repeated for the other scaling factors shown in Figures 9.10 and 9.11. Since the restrictions $\lambda^{(j)} = 0$ and $\lambda^{(j)} = 1$ give models with the same number of free parameters, the choice between them depends entirely on the log-likelihood found when fitting the model. However, the difference between $\lambda^{(j)} = 0$ and $\lambda^{(j)} = 1$ is the difference between a model which allows mortality rates to change with time and a static model of mortality ($\lambda^{(j)} = 0 \; \forall j$). We therefore find that, in very small schemes it is almost impossible to say whether or not mortality rates are changing, let alone anything about the rate they are changing.

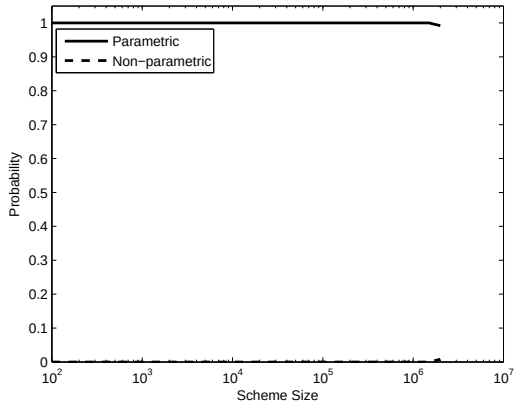
Looking at Figure 9.10b again, we see that for larger schemes, with around 10,000 to 100,000 members, the relative model has a clear preference for setting $\lambda^{(1)} = 1$ for men, which is preferred in almost all simulations for schemes with around 200,000 members. This pattern is also true for the majority of the scaling factors shown in Figures 9.10 and 9.11. The implication of this is that, although there is sufficient evidence to suggest mortality is improving in these larger schemes (unlike the smaller schemes discussed above), there is not enough data to quantify any differences in this improvement between the scheme and the national population. This supports the use of projection methods based on the national population for the majority of pension schemes in the UK. It also makes it unlikely that we can detect trend basis risk between the scheme and the national



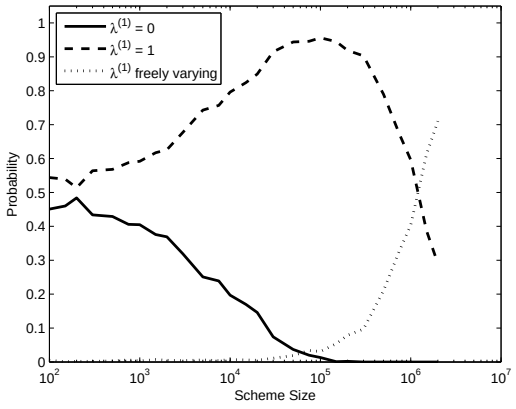
(A) $\alpha_x^{(\Delta)}$ (B) $\kappa_t^{(1)}$ (C) $\kappa_t^{(2)}$

(D) $\kappa_t^{(3)}$ (E) γ_y

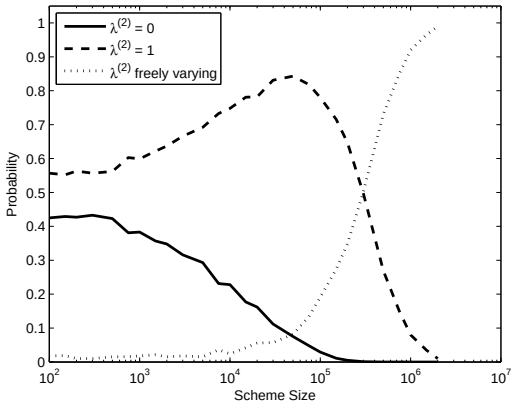
FIGURE 9.10: Restrictions placed on the relative model for different volumes of male data



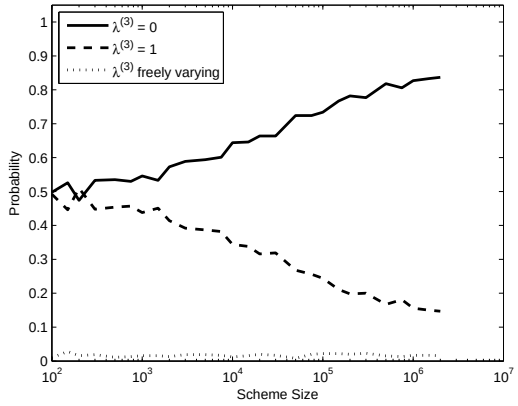
(A) $\alpha_x^{(\Delta)}$



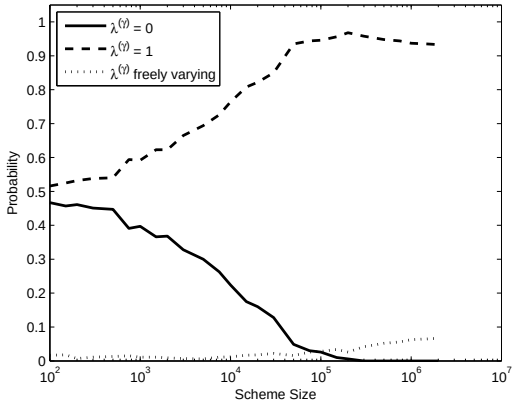
(B) $\kappa_t^{(1)}$



(C) $\kappa_t^{(2)}$



(D) $\kappa_t^{(3)}$



(E) γ_y

FIGURE 9.11: Restrictions placed on the relative model for different volumes of female data

population for schemes with fewer than 100,000 members of each sex.

Only in the very largest schemes, with over one million members of each sex, do we find that there is sufficient data to estimate unrestricted $\lambda^{(j)}$, as illustrated by the preference for a freely varying $\lambda^{(1)}$ for men for schemes with around two million members in Figure 9.10b. Therefore, it is only for these very large schemes that we can quantify any difference in the evolution of mortality rates between a pension scheme and the national population, i.e., any trend basis risk, although the results of Section 9.5 indicate that, even when this is allowed for, the impact on annuity values is likely to be quite limited, especially when considered in the context of the other mortality and longevity risks in the scheme. This is investigated further in Chapter 10 .

In summary, we find that, for datasets that are the same size as a typical UK pension scheme, there is insufficient data to make more than a few simple adjustments to reflect level basis risk. For most practical circumstances, we would therefore be unable to quantify any trend basis risk in a pension scheme. Given that trend basis risk is often given as a key concern for why pension schemes are reluctant to use index based hedging instruments to manage their longevity risk and, instead, prefer bespoke arrangements, we believe that much of this trepidation is misplaced, as we now discuss.

9.7 Discussion: Basis risk in pension schemes

There has been a lot of work regarding the quantification of basis risk between different populations, most notably in Plat (2009b), Salhi and Loisel (2009), Li and Hardy (2011), Coughlan et al. (2011), Cairns et al. (2013) and Li et al. (2013). The analysis of this risk has also motivated many of the multi-population mortality models that have recently been proposed, such as those of Dowd et al. (2011b), Cairns et al. (2011a), Zhou et al. (2014) and Chapter 8. However, much of this work to date is not directly relevant to the situation faced by many UK pension schemes when assessing and trying to manage their longevity risk.

Partly, this is because the populations being considered in these studies are far larger in terms of the size of the exposures to risk than that of a typical (or, indeed, even a very large) UK pension scheme. This enables the authors of these studies to adopt a "general-to-specific" approach when analysing trend basis risk: first mortality models are

fitted separately to the different populations under investigation and then any dependence between the period or cohort parameters is analysed. This approach is exemplified by the study of Li et al. (2013), which statistically determined whether or not to simplify a model by using the same sets of parameters for different populations (which is a very specific form of dependence). Such an approach therefore starts from the assumption that mortality rates will have different patterns of evolution in different populations, and then looks for evidence of similarities.

Such an approach is entirely reasonable when looking at large populations where there is sufficient data to estimate sophisticated mortality models in each population under investigation. However, this is not the situation in which most pension schemes find themselves. Instead, with relatively little data, it is necessary for them to adopt a "specific-to-general" approach, such as that underlying the relative model proposed in this study. As there is insufficient data to estimate many sub-population-specific parameters robustly, a specific-to-general methodology starts from the assumption that mortality rates in the sub-population evolve in the same fashion as those in the reference population and then looks for evidence of differences between the two. This approach naturally leads to more parsimonious models, which are therefore likely to be more robust. However, it is less likely to overturn the null hypothesis of no trend basis risk, especially when parameter uncertainty and model risk are included in any analysis.

Our findings suggest that large volumes of data (in terms of both the size of the exposures to risk and the period range of the data) are required to overturn the null hypothesis of no trend basis risk, especially when parameter uncertainty and model risk are included in the analysis. For the full SAPS dataset, the simple relative model we have proposed achieves relatively good and parsimonious fits to the data for both men and women, as shown in Section 9.4. Furthermore, for the smaller datasets more typical of UK pension schemes, even simpler models which fix the scaling factors in the model are preferred, as shown in Section 9.6. This is consistent with the results of Haberman et al. (2014), which found that it is only possible to quantify basis risk for very large schemes.

In addition, in order to estimate the more complicated multivariate time series processes used in many of the general-to-specific models we need longer periods of data than a typical pension scheme has. For instance, to estimate the cointegration-based models of Salhi and Loisel (2009) and Chapter 8 requires several decades of mortality data, which is usually far in excess of what a pension scheme will have itself. Similarly, Haberman et al. (2014) found that eight years or more of data is required for the quantification of

basis risk, even for very large pension schemes. Specific-to-general models, however, do not require such long data ranges, as they start from the assumption that information about the reference population can be used to fill in gaps in the data if required.

However, Section 9.5 shows that projections from the relative model have many of the features we would expect from models which use more complicated time series processes, when appropriate allowance is made for parameter uncertainty and model risk, despite there being no genuine trend basis risk using the relative approach. This implies that it may be impossible to distinguish between genuine trend basis risk and the effects of parameter uncertainty and model risk in practice. Indeed, it is noticeable that few of the studies to date which have investigated basis risk allow for parameter uncertainty and model risk, and so the findings of these studies potentially wrongly attribute differences in historical improvements in mortality between different populations to basis risk and, thus, overstate its importance.

We find that for most UK pension schemes, the existence or not of trend basis risk between the scheme and the UK population is of little practical relevance. The scheme will never have sufficient information to be able to say with confidence that the improvements in mortality it experiences are significantly different from that in the reference population, as any such differences will be overwhelmed by the other sources of risk and uncertainty present in the scheme.

This is not to dispute that basis risk can exist between different countries or amongst highly distinct sub-populations of a reference population. Indeed, there are good reasons to suggest that it does and that there is sufficient data to estimate it reliably using a general-to-specific approach as in previous studies. For instance, many studies (for instance in Li and Hardy (2011) and Chapter 8) investigate differences between the evolution of mortality rates in different countries. However, populations in different countries may have different diets, lifestyles and access to healthcare, and so would be expected to have different patterns of evolution in mortality rates. Other studies, such as in Villegas and Haberman (2014) consider the differences in the evolution of mortality rates between highly selective sub-populations of a country (for instance, based on deprivation). The sub-populations in these studies have, therefore, been constructed in such as fashion as to maximise the likelihood of observing different patterns in the evolution of mortality rates.[18]

[18] As well as being a highly selected sub-population of the UK population, the data for CMI Assured Lives has also varied considerably in the socio-economic makeup of the relevant population over the period of the data due to changes in the UK annuity market. As this dataset was used in Cairns et al.

Nor do we argue that the evolution of mortality rates in a pension scheme *is* the same as in the reference population. It may be true that for very large schemes, we may have sufficient data to be able to detect trend basis risk (even when allowing for parameter uncertainty and model risk) if there is quite a large difference in the evolution of mortality rates between the two populations.

However, a pension scheme, whose only membership requirement was employment with a particular company, would be expected to be more similar to the national population or differ only due to persistent selection effects which affect the level of mortality rates but not how mortality rates evolve with time. In order to have sufficient data to reject the assumption that the evolution of mortality rates in the pension scheme is the same as in the national population, the scheme must be very large (such as being the pension scheme for a large and long-established national company) and so entry to such schemes is likely to be relatively unselective. Therefore, these schemes are more likely to represent a fair cross section of the UK population. Consequently, the circumstances where we have enough data to quantify basis risk (for example, the pension scheme of a large, national employer) are also the circumstances when basis risk is least likely to be important. Consequently, in most practical situations, we will never have sufficient data to tell the difference and therefore an assumption of no difference between the evolution of mortality rates in the national population and the pension scheme is both practical and parsimonious.

The practical implications of these results are important for the development of any market in longevity hedging. As trend basis risk is unlikely to be important enough to be statistically significant, it is also unlikely to be financially significant. If longevity risk is felt to be important, hedging can be achieved by use of standardised instruments based on projected changes in mortality rates in a reference population, making adjustments to reflect the level of mortality observed in the pension scheme. Concerns that the trend basis risk will make such hedges ineffective, such as those raised against the EIB longevity bond (see Blake et al. (2006)), should be regarded as secondary compared with the other risks a pension scheme faces, such as idiosyncratic mortality risk. Bespoke products, such as longevity swaps tailored to the characteristics of the pension scheme, should be regarded primarily as vehicles for hedging and transferring these other risks, rather than any trend basis risk for the scheme, and their cost effectiveness judged accordingly, as

(2011a), Dowd et al. (2011b) and Cairns et al. (2013), it is therefore unclear whether any difference in the evolution of mortality detected by these studies is the result of genuine trend basis risk or simply a result of the changing composition of the dataset.

discussed in Chapter 10 .

9.8 Conclusions

In conclusion, in this study we present a relative model for mortality in a sub-population, which models the mortality rates observed in a small population relative to those observed in a larger reference population. Such a model has the advantages of being more parsimonious compared with the approach of fitting separate mortality models for both populations, which has been adopted in many multi-population mortality studies, and so is better suited to situations where there is little data for the sub-population.

We then apply the relative model to investigate the mortality rates observed in the SAPS study of UK pension schemes. We find that this simple model is sufficient to achieve a good and parsimonious fit to the available data and reasonable projections of mortality rates. Specifically, we find that, in aggregate, members of UK occupational pension schemes generally experience lower levels of mortality rates than the national population, which are also improving at a faster rate than those in the national population. However, we find relatively high levels of uncertainty in estimating the parameters even in this simple model and that the data is insufficient to uniformly prefer one model over any other. Furthermore, when we apply the relative modelling approach to sub-populations which are smaller than the SAPS population, and closer in size to those of typical UK pension schemes, we find that the modelling approach prefers very simple, highly restricted models, which do not allow for any difference in the evolution of mortality between the reference and sub-populations.

These considerations lead us to the belief that the analysis of trend basis risk, which requires more sophisticated models than the relative model proposed, is not possible with the datasets realistically available for most pension schemes. We find that, in pension scheme sized datasets, we will never have sufficient evidence to determine whether there is any difference in the evolution of mortality rates in the sub-population compared with the reference population when the other risks present are properly accounted for. Therefore, we believe that an assumption of no difference in the evolution of mortality rates between the two populations is practical and parsimonious. Consequently, we conclude that concerns regarding trend basis risk in the development of the market for longevity hedging and risk management tools for pension schemes are misplaced.

9.A Summary of SAPS data

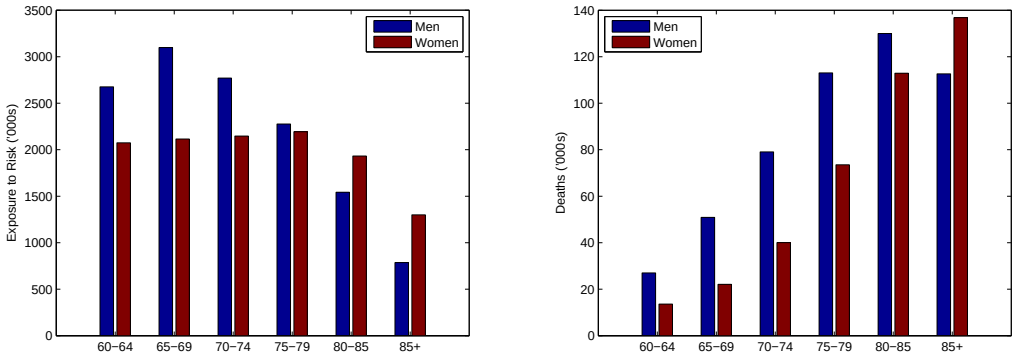
We are indebted to the CMI for kindly providing death counts and exposures, weighted by individual lives, for the SAPS population for the period 2000 to 2011 and ages 60 to 90. These relate to all pensioners in the surveyed pension schemes, and so include people receiving benefits after retiring at normal retirement age, those who retired early or in ill-health, and those in receipt of spousal benefits. It is likely that some of these sub-populations will have different mortality characteristics, especially those retiring in ill-health. However, such cases represent a relatively small proportion of the SAPS data and are unlikely to materially impact our results.

Large pension schemes in the UK submit their mortality experience to the SAPS study following completion of a triennial funding valuation. Therefore, each submission is in respect of data with a considerable time delay, e.g., data submitted on 30 June 2013 may result from a funding valuation with an effective date of 31 December 2011 (due to the time taken to perform the valuation) and cover the period 1 January 2009 to 31 December 2011. Consequently, the last few years of the SAPS data only reflects a partial submission to date of the mortality experience of the schemes which will, ultimately, submit data to the study. However, we have no reason to believe that the schemes that have submitted to date are an unrepresentative sub-sample of the SAPS population, and so do not believe this biases our results.

Similarly, there are fewer submissions for the earliest years of the SAPS data. Unlike the most recent years, the missing data for this period will never be received by the CMI. Therefore, we only have data we consider complete for roughly the period 2004 to 2008.[19]

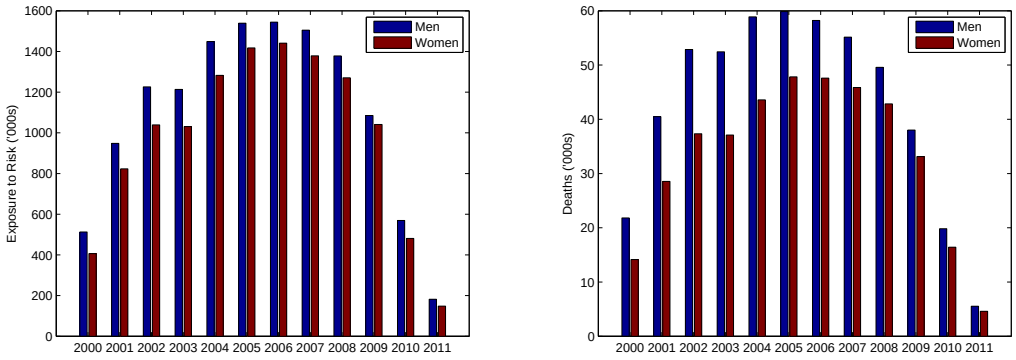
Figures 9.12 and 9.13 summarise the patterns of deaths and exposures for men and women across age and time.

[19] However, we note that Continuous Mortality Investigation (2014b) and Continuous Mortality Investigation (2014c) have been published subsequently to us obtaining the data used in this study from the CMI. These working papers included new data in respect of the SAPS study for 2012 and 2013, respectively, along with revisions to the data for years prior to 2012 caused by new pension schemes submitting data to the study. In the interests of avoiding data errors caused by merging multiple sources of data, we have not combined this new data with that provided previously by the CMI and, therefore, it has not been included in this study. However, we have investigated the impact the new data would have on our findings if it were included, and are satisfied that it would not affect our results materially.



(A) Exposure to risk by age band (B) Death count by age band

FIGURE 9.12: Exposures to risk and death counts in the SAPS dataset by age



(A) Exposure to risk by year (B) Death count by year

FIGURE 9.13: Exposures to risk and death counts in the SAPS dataset by year

9.B Identifiability in the relative model

In Chapters 3 and 4, we discussed the identifiability issues in age/period and age/period/cohort mortality models, respectively. In particular, we find that almost all APC mortality models possess "invariant" transformations, i.e., transformations of the parameters of the model which leave the fitted mortality rates unchanged. In order to find a unique set of parameters, we impose a set of identifiability constraints on them. Typically, these are chosen so that we can assign our desired interpretation of the demographic significance to the parameters in question. However, because this interpretation is subjective, it is important that our choice of identifiability constraints does not have any impact on any observable quantities. For instance, we discuss in Chapters 3 and 4 how to ensure that projected mortality rates are independent of the choice of identifiability constraints.

The relative model in Equation 9.3 does not possess any additional identifiability issues in and of itself, once the parameters from the reference population are known. However, due to the relative structure, transformations of the parameters in the reference population model will have knock-on effects for those in the relative model. It is important therefore that invariant transformations of the reference model are also invariant for the relative model, so that our choice of identifiability constraints for the reference population does not affect the suitability of the relative model. This requirement will determine both the nature of the set of deterministic functions of year of birth, X_y in Equation 9.3, and the nature of any parametric simplification imposed upon $\alpha_x^{(\Delta)}$, i.e., if $\alpha_x^{(\Delta)}$ is restricted to be a linear combination of a set of basis functions

$$\alpha_x^{(\Delta)} = \sum_{i=1}^{n} \alpha^{(i)} g^{(i)}(x)$$

then the nature of the basis functions, $g^{(i)}(x)$, will be determined by the identifiability issues present in the model. We, therefore, consider each of the different forms that the invariant transformations of the reference model can take in turn, in order to ensure that they will not affect the relative model.

First, the sensitivities in the relative model trivially do not depend upon the normalisation scheme of the age/period terms in the reference model. Normalisation schemes are imposed by using a transformation of the form

$$\{\hat{f}^{(R,i)}(x), \hat{\kappa}_t^{(R,i)}\} = \left\{ \frac{1}{a^{(i)}} f^{(R,i)}(x), a^{(i)} \kappa_t^{(R,I)} \right\}$$

and so it is obvious that $\Lambda^{(i)} \hat{f}^{(R,i)}(x) \hat{\kappa}_t^{(R,i)} = \Lambda^{(i)} f^{(R,i)}(x) \kappa_t^{(R,i)}$.

Second, we know from Chapter 3 that all APC models are invariant under the transformation

$$\{\hat{\alpha}_x^{(R)}, \hat{f}^{(R,i)}(x), \kappa_t^{(R,i)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - a^{(i)} f^{(R,i)}(x), f^{(i)}(x), \kappa_t^{(R,i)} + a^{(i)}, \gamma_y^{(R)}\} \quad (9.12)$$

i.e., the model using the transformed parameter set gives exactly the same fitted mortality rates. This allows us to impose the "level" of the period functions, $\kappa_t^{(R,i)}$, via the identifiability constraints, such as imposing $\sum_t \kappa_t^{(R,i)} = 0$ or $\kappa_T^{(R,i)} = 0$. However, such a set of identifiability constraints is arbitrary, and so should not have any consequences for our relative modelling approach.

Accordingly, we require that our relative model in Equation 9.3 is also invariant if the transformed parameters are used for the reference population. In order to ensure this, we require that Equation 9.3 is invariant under the transformation

$$\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - a^{(i)}\Lambda^{(i)} f^{(R,i)}(x) \tag{9.13}$$

This transformation can be accommodated without $\alpha_x^{(\Delta)}$ fundamentally changing form if

1. $\alpha_x^{(\Delta)}$ is non-parametric, as in the original specification in Equation 9.3; or
2. if $\alpha_x^{(\Delta)}$ is restricted to be of parametric form, then $\alpha_x^{(\Delta)} = \sum_{i=1}^{N} \alpha^{(i)} f^{(i)}(x) + \sum_{i=N+1}^{n} \alpha^{(i)} g^{(i)}(x)$, i.e., the age functions in the reference model form a subset of the basis functions, $g^{(i)}(x)$.

As an example, consider the case where our model for the reference population is the "classic APC" model of Hobcraft et al. (1982)

$$\begin{aligned}\ln\left(\mu_{x,t}^{(R)}\right) &= \alpha_x^{(R)} + \kappa_t^{(R)} + \gamma_{t-x}^{(R)} \\ R_{x,t} &= \alpha_x^{(\Delta)} + \Lambda^{(1)}\kappa_t^{(R)} + \Lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \nu X_{t-x}\end{aligned}$$

The classic APC model is invariant under the transformation

$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - a, \kappa_t^{(R)} + a, \gamma_y^{(R)}\}$$

i.e., $\hat{\mu}_{x,t}^{(R)} = \mu_{x,t}^{(R)}$. Substituting the transformed parameters into the relative model gives

$$\begin{aligned}\hat{R}_{x,t} &= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}\hat{\kappa}_t^{(R)} + \hat{\Lambda}^{(\gamma)}\hat{\gamma}_{t-x}^{(R)} + \hat{\nu}X_{t-x} \\ &= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}(\kappa_t^{(R)} + a) + \hat{\Lambda}^{(\gamma)}\gamma_{t-x}^{(R)} + \hat{\nu}X_{t-x}\end{aligned}$$

In order to ensure $\hat{R}_{x,t} = R_{x,t}$, we must have $\hat{\Lambda}^{(1)} = \Lambda^{(1)}$, $\hat{\nu} = \nu$ and $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - a\Lambda^{(1)}$. The requirement that $\hat{\alpha}_x^{(\Delta)}$ is of the same form as $\alpha_x^{(\Delta)}$ implies that any parametric simplification for $\alpha_x^{(\Delta)}$ must be of the form $\alpha_x^{(\Delta)} = \alpha^{(1)} + \sum_{j=2}^{n} \alpha^{(i)} g^{(i)}(x)$, i.e., it has a constant basis function, $g^{(1)}(x) = 1$, in order that the relative model does not change if the levels of the period functions are transformed.

Third, the values of $\Lambda^{(i)}$ depend upon the precise definition of the age functions in the reference model. "Equivalent" models for the reference population, which use different definitions for the age functions but give identical fitted mortality rates, will give different

values of $\Lambda^{(i)}$. To see this, consider a reference model of the form[20]

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \kappa_t^{(R,1)} + (x-\bar{x})\kappa_t^{(R,2)} + \gamma_{t-x}^{(R)}$$
$$R_{x,t} = \alpha_x^{(\Delta)} + \Lambda^{(1)}\kappa_t^{(R,1)} + \Lambda^{(2)}(x-\bar{x})\kappa_t^{(R,2)} + \Lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \nu X_{t-x}$$

The model for the reference population is equivalent to a model of the form

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \hat{\kappa}_t^{(R,1)} + x\hat{\kappa}_t^{(R,2)} + \gamma_{t-x}^{(R)}$$

with $\hat{\kappa}_t^{(R,1)} = \kappa_t^{(R,1)} - \bar{x}\kappa_t^{(R,2)}$ and $\hat{\kappa}_t^{(R,2)} = \kappa_t^{(R,2)}$. The corresponding relative model in this case is

$$\hat{R}_{x,t} = \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}\hat{\kappa}_t^{(R,1)} + \hat{\Lambda}^{(2)}x\hat{\kappa}_t^{(R,2)} + \Lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \hat{\nu}X_{t-x}$$

However, in this situation, we would find that $\hat{\Lambda}^{(2)}x = \Lambda^{(2)}(x-\bar{x}) + \Lambda^{(1)}\bar{x}$ in order to give the same fitted mortality rates for both reference models. If so, the relationship between the two would be a function of age, x, which contradicts the assumption that the scaling factors are constants independent of age. Consequently, we find that the values of the scaling factors and the fit provided by the relative model will depend on the specifics of the age functions in the reference model and will differ between equivalent models.

Finally, identifiability under transformations of the cohort parameters is not as straightforward. From Chapter 4, we found that APC models may have unidentifiable trends which are allocated between the age/period and cohort terms by the identifiability constraints. Invariance of the mortality rates in the relative model to a different allocation of these trends in the reference model depends upon the deterministic regressors, X_y, we added to the relative model in Equation 9.3, and the form of any parametric simplification of $\alpha_x^{(\Delta)}$. This is illustrated by the following example.

Consider the example of the classic APC model for the reference population again. In addition to the transformation above, the classic APC model is also invariant under the following two transformations involving the cohort parameters

$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} - b, \kappa_t^{(R)}, \gamma_y^{(R)} + b\}$$
$$\{\hat{\alpha}_x^{(R)}, \hat{\kappa}_t^{(R)}, \hat{\gamma}_y^{(R)}\} = \{\alpha_x^{(R)} + c(x-\bar{x}), \kappa_t^{(R)} - c(t-\bar{t}), \gamma_y^{(R)} + c(y-\bar{y})\}$$

[20] We call this model the "reduced Plat" model, since it was suggested in Plat (2009a) as being a reduced form of the model tested in that paper that might be more suitable for high ages. This model can also be thought of as an extension to model M6 in Cairns et al. (2009), with a static age function, or as an extension to the "CBDX" model discussed in Chapter 3 with a cohort term.

where a bar denotes the arithmetic mean of the variable over the relevant data range.[21] Invariance of the relative model under the first of these transformations requires $\hat{\Lambda}^{(\gamma)} = \Lambda^{(\gamma)}$ and $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - b\Lambda^{(\gamma)}$, and therefore that any parametric restriction placed upon $\alpha_x^{(\Delta)}$ must have a constant basis function, $g^{(1)}(x) = 1$, as discussed above in respect of the level of $\kappa_t^{(R)}$.

However, substituting the transformed parameters from the second transformation in Equation 9.3, we find

$$\begin{aligned}\hat{R}_{x,t} &= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}\hat{\kappa}_t^{(R)} + \hat{\Lambda}^{(\gamma)}\hat{\gamma}_{t-x}^{(R)} + \hat{\nu}X_{t-x} \\ &= \hat{\alpha}_x^{(\Delta)} + \hat{\Lambda}^{(1)}(\kappa_t^{(R)} - c(t-\bar{t})) + \hat{\Lambda}^{(\gamma)}(\gamma_{t-x}^{(R)} + c((t-\bar{t}) - (x-\bar{x}))) + \hat{\nu}X_{t-x}\end{aligned}$$

In order to have $\hat{R}_{x,t} = R_{x,t}$, we require

- $\hat{\Lambda}^{(j)} = \Lambda^{(j)}$, i.e., that our sensitivities do not change from one set of identifiability conditions to any other;
- $\hat{\nu}X_y = \nu X_y - c(\lambda^{(\gamma)} - \lambda^{(1)})(y - \bar{y})$, i.e., we can add terms linear in year of birth to the deterministic term without it fundamentally changing form, and therefore that our deterministic regressors contain a linear trend in year of birth; and
- $\hat{\alpha}_x^{(\Delta)} = \alpha_x^{(\Delta)} - c\lambda^{(1)}(x - \bar{x})$, i.e., we can add linear functions to any parametric form for $\alpha_x^{(\Delta)}$ without it fundamentally changing form, and therefore that it must be either non-parametric or have a linear function of age, $g^{(2)}(x) = x - \bar{x}$, amongst the basis functions used in any parametric restriction.

In addition to the identifiability issues discussed here, it is also important that any parametric simplification for $\alpha_x^{(\Delta)}$ consists of more than one, constant term. As discussed in Tuljapurkar and Edwards (2009), multiple terms in $\alpha_x^{(\Delta)}$ allow higher moments of the observable distribution of deaths in the sub-population (such as the variance of age at death) to be captured by the relative model, as well as the difference in life expectancy between the two populations. These higher moments are important in the allowance for idiosyncratic risk in the sub-population, which is likely to be important in many circumstances, such as those discussed in Chapter 10 .

We also see from the analysis above that the form of our deterministic regressors, X_y, will depend upon the mortality model being used for the reference population. From Chapter 4, if the model for the reference population contains age functions which span the

[21] e.g., $\bar{x} = \frac{1}{X}\sum_x x = 0.5(X+1)$ and similarly for $\bar{t}$ and $\bar{y}$.

polynomials to order p, then there will be unidentified polynomial trends in the cohort parameters of order $p+1$. We must therefore ensure that the deterministic regressors in Equation 9.3 span the polynomials to order $p+1$ and that any parametric simplification for the age function, $\alpha_x^{(\Delta)}$, also contains a basis function of the form $g^{(i)}(x) = x^{p+1}$.

For the classic APC model and the models constructed by the general procedure in Section 9.4.1 and Appendix 9.C, $p = 0$ and therefore we require that the deterministic regressors and age function are, at least, of linear order. Similarly, for the reduced Plat model, $p = 1$, and therefore we would require that the deterministic regressors are at least of quadratic order.

In summary, the identifiability issues present in APC mortality models and discussed in Chapters 3 and 4 have important consequences for the relative mortality modelling approach used in this study. Most importantly, we require an additional νX_y term in the model and must be careful when specifying any parametric simplification for $\alpha_x^{(\Delta)}$, in order to ensure that our results do not depend on the arbitrary identifiability constraints we impose on the reference model. In the context of the reference model used in this study, described in Section 9.4.1, this means that we need the term

$$\nu X_y = \nu_1 (y - \bar{y})$$

in Equation 9.3, and any parametric simplification of $\alpha_x^{(\Delta)}$ must be of the form

$$\begin{aligned}\alpha_x^{(\Delta)} &= \begin{pmatrix}\alpha^{(1)}, & \alpha^{(2)}, & \alpha^{(3)}, & \alpha^{(4)}\end{pmatrix} \begin{pmatrix} f^{(1)}(x) \\ f^{(2)}(x) \\ f^{(3)}(x) \\ (x - \bar{x}) \end{pmatrix} \\ &= \sum_{i=1}^{N+1} \alpha^{(i)} \tilde{f}^{(i)}(x)\end{aligned}$$

where $f^{(i)}(x)$ are the parametric age functions in the reference model, described in Table 9.1.

9.C Models constructed by the "general procedure" for the UK

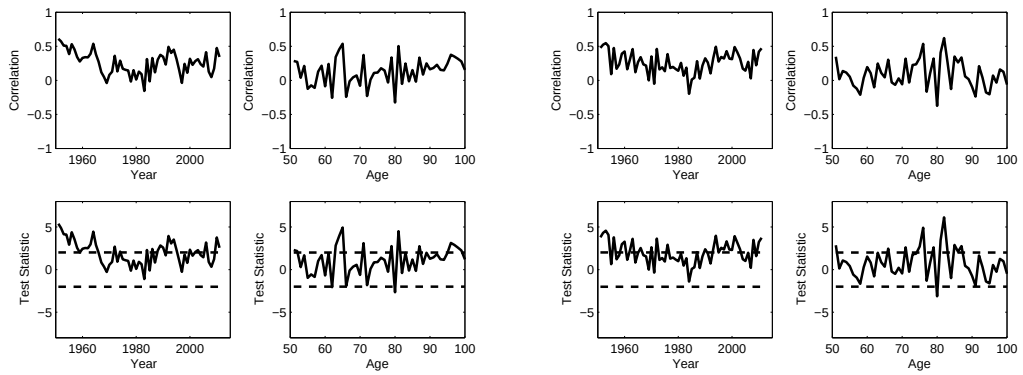
In Chapter 5, a "general procedure" for constructing mortality models tailored to the specific features of individual datasets was proposed. In outline, this

- starts from a simple static mortality model with a non-parametric static age function;
- sequentially adds age/period terms to the model to detect and capture the age/period structure in the data:
  - structure is detected by adding a non-parametric age/period term which will identify the feature explaining the largest proportion of the remaining structure in the data;
  - then this term is simplified into a parametric form which identifies the same feature more parsimoniously and with greater demographic significance;
  - then the statistical significance and robustness of the term is tested;
- finally adds a cohort term once all age/period structure has been captured by the model;
- tests the standardised deviance residuals of the model for any remaining structure, independence, and normality.

This procedure was applied to data from the Human Mortality Database (2014) for men and women in the UK for ages 50 to 100 and years 1950 to 2011 in order to construct mortality models capable of capturing all the relevant information in the data and therefore allowing it to be projected appropriately.

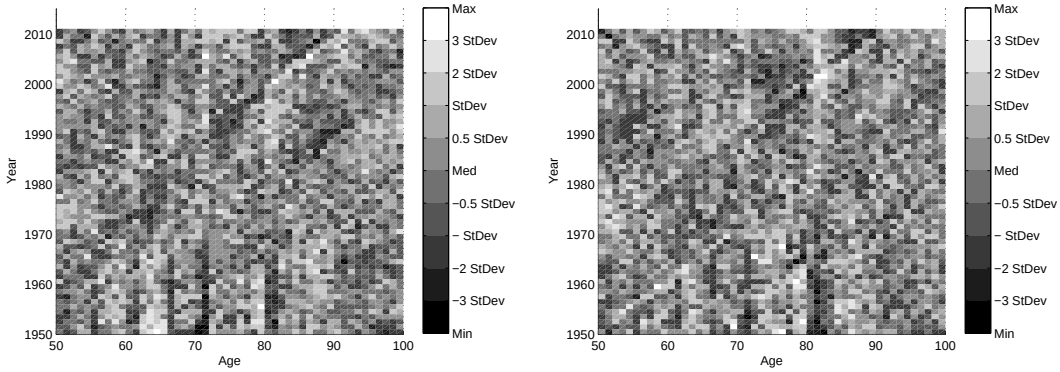
A brief description of the terms in the models and their demographic significance is given in Table 9.1. A fuller list of the parametric age functions in the "toolkit" developed as part of the general procedure is given in the Appendix of Chapter 5.

As discussed in Section 9.4, we also require additional identifiability constraints in order to obtain a unique set of parameters when fitting the model to data. These are given in Section 9.4 and have been chosen to aid comparability between the models for the reference population and the relative model in Equation 9.3.



(A) UK men (B) UK women

FIGURE 9.14: Correlations for sequential years and ages of the residuals from fitting the model developed by the general procedure to data for the UK



(A) UK men (B) UK women

FIGURE 9.15: Heat maps of the residuals from fitting the model developed by the general procedure to data for the UK

When fitting the final models, we obtain the parameters shown in Figures 9.1 and 9.2. These models have BICs of -1.95×10^4 and -1.99×10^4 for men and women, respectively, with 345 and 346 free parameters.[22] We also test the standardised deviance residuals from fitting the model as part of the general procedure. The moments of the residuals and a Jarque-Bera test of their normality is given in Table 9.6. We can see that the residuals are close to normal, although they are slightly leptokurtic for both datasets and therefore fail the relevant Jarque-Bera tests for normality at the 5% level (p-values of 2.8% for men and 0.2% for women). We also see from Figures 9.14a and 9.14b that there appears to be relatively little correlation structure over consecutive ages, although the residuals show significant autocorrelations during the early part of the data range, which diminishes towards the end of the period of the data.

[22] For comparison, the Lee and Carter (1992) model fitted to the same data obtains BICs of -2.71×10^4 and -2.69×10^4 with 161 free parameters for both populations.

	Residual mean	Standard deviation	Residual skewness	Residual kurtosis	Jarque-Bera statistic
Men	-0.01	0.94	-0.01	3.19	4.76
Women	-0.01	0.94	-0.01	3.32	13.97

TABLE 9.6: Moments of the residuals from fitting the model developed by the general procedure to data for the UK for men and women in the UK

The heat maps for the residuals shown in Figure 9.15 indicate that the residuals for both sexes in the UK have very little remaining structure in them. There is possibly some remaining structure around age 80 for both men and women, although this appears to be specific to only a few neighbouring years and therefore it is difficult to add an age/period term to capture this without overfitting the models

Chapter 10

Transferring Risk in Pension Schemes via Bespoke Longevity Swaps

10.1 Introduction

The pensions de-risking industry has grown enormously in recent years, especially in the UK which has pioneered many of the de-risking techniques which have since become international. The sponsors and trustees of pension schemes[1] in the UK have increasingly looked to both reduce or transfer the risks in providing defined benefit pensions to scheme members. This has included reviewing schemes' investment strategies to match the timing and nature of the projected cashflows (called "liability-driven investment" or LDI) and limiting the accrual of benefits to new and existing members of the scheme. Indeed, the majority of private sector pension schemes in the UK are now closed to the future accrual of benefits, meaning that they are now solely responsible for managing the run-off of the legacy benefits for members. In more recent years, the focus of this de-risking has been to transfer the financial and demographic risks of the scheme to third parties, either by a "buy-out" or a "buy-in".

Longevity swaps were developed to hedge and transfer mortality and longevity risks directly, without reference to the other investment and financial risks present in the scheme. The market for bespoke longevity swaps - those defined with reference to the

[1] In this chapter, we refer to "pension schemes" which administer the provision of retirement benefits defined in terms of salary and service to members. We would draw a semantic distinction between a "pension scheme" and a "pension plan", which we would use as a more general term for any defined benefit or defined contribution pension arrangement provided on either a group or an individual basis.

specific characteristics of the pension scheme membership - has grown exponentially to over £50bn in the UK, which currently leads the world in this area.[2]

In this chapter, we present a modelling framework suitable for assessing the various mortality and longevity risks within a stylised pension scheme and the effectiveness of a bespoke longevity swap in reducing the risks faced by the scheme. In particular, we focus on the possible interactions between the different risk factors that influence mortality rates, which are often overlooked in existing studies. Since this is the first study to look at these issues in detail, some of the allowances we make for these risk factors are approximate in nature, and are based on our professional experience of pensions consultancy in the UK, advising on buy-ins, buy-outs and longevity swaps, rather than established stochastic models. However, we are confident that the impact of these allowances is broadly reasonable and consistent with our practical experience, but are aware that further research is required.

In order to achieve a comprehensive analysis of these risks, we distinguish between "longevity risk", referring to systematic mortality-related risks in the pension scheme (i.e., those relating to nation-wide and scheme-wide populations), and "mortality risk", is referring to those mortality-related risks which are specific to the individual members of the scheme. We do this by first investigating the systematic longevity risk in the national population, before assessing the scheme-specific longevity basis risks present and, finally, making appropriate allowance for individual characteristics and the idiosyncratic mortality risk. We apply this analysis to stylised pension scheme data, which has been generated to incorporate many of the features observed in real pension schemes.

The chapter is structured as follows. First, in Section 10.2, we review the markets for pension scheme de-risking in the UK, and, in particular, the market for bespoke longevity swaps. In Section 10.3, we discuss the data for the stylised pension scheme used in this study. Section 10.4 considers the modelling approach used to quantify the various mortality and longevity risks present in this illustrative scheme. In Section 10.5, we compare the future cashflows calculated using the assumptions for mortality rates which are often made in practice in the UK with those which would be projected as a "best estimate" from the stochastic mortality models we use to assess mortality and longevity risks and provide a bridge between the two. Then, in Section 10.6, we measure the contribution of the different stochastic mortality and longevity risks for the scheme, with a particular emphasis on the cost effectiveness of a bespoke longevity swap in managing

[2]For instance, see Hymans Robertson (2015).

these risks. Finally, in Section 10.7 we discuss our findings and their implications for the further development of the longevity swap market both in the UK and internationally.

10.2 Longevity swaps

Over the past decade, as the perceived importance of longevity risk has grown, a number of new tools have emerged to allow pension schemes to manage this risk.[3] Originally, a pension scheme wishing to transfer longevity risk to an insurer would have to do so via a "buy-out". This would involve purchasing either immediate or deferred annuities in the name of each scheme member matching the members' accrued benefits within the scheme. Thus, the scheme would fully transfer all of the assets and liabilities of the scheme to the insurer and discharge its obligation completely. Typically, this was very expensive, in part due to limited competition in the market for buy-outs, which meant it was usually only done when the scheme was wound up following the insolvency of the sponsoring employer. However, the emergence of new life insurers specialising in buy-outs in the mid-2000s brought the cost down to some extent and, so, made buy-outs feasible during corporate transactions to extinguish the ongoing obligation of the acquiring company to the pension scheme.

One major innovation in the pensions risk-management market was the development of pension "buy-ins" as an alternative to the full risk transfer of a buy-out. A buy-in involves the scheme purchasing an insurance contract which is tailored to exactly replicate the benefits payable to a subset of the scheme members (usually pensioners). Unlike a buy-out, the insurance contract is an asset of the scheme rather than of the individual scheme members. Payments from the buy-in contract are not earmarked for the specific members covered by the contract and, in the event of insolvency of the insurance company, the scheme retains the obligation to provide benefits to the covered members. Therefore, a buy-in represents an investment decision to purchase a (perfect) hedging instrument for the future benefit payments rather than a full transfer of the risk to another party.

Unlike a buy-out, the scheme can purchase a buy-in for a subset of scheme members without adversely affecting those not covered by the contract (since the preferential treatment of one section of scheme members is not permitted in the UK). This enables buy-ins in respect of only the pensioner members of the scheme, rather than deferred members,

[3] See Blake et al. (2013) for a more detailed survey of developments in the "new life market".

which substantially reduces the cost of a buy-in arrangement.[4] Furthermore, in a buy-in, the scheme remains liable for members' benefits in the event of insolvency of the life insurer. In practice, however, most life insurers are considerably more creditworthy than the sponsoring employers of the scheme purchasing a buy-in and policyholders receive a high level of compensation under the Financial Services Compensation Scheme in the unlikely event of insurer insolvency, and so credit risk is considered negligible in the UK.

Since a buy-in contract matches the benefit structure of the covered members exactly, it mitigates the investment and inflation risk as well as the demographic risks, such as longevity risk, in respect of these members. However, the scheme retains financial and demographic risks for non-pensioner members, which may be desirable if it feels it can profit from the upside of these risks or they are too expensive to transfer immediately.

In contrast, a "longevity swap" represents a pure transfer of longevity risk, with no mitigation of investment or inflation risks.[5] As a consequence, longevity swaps are usually less expensive than buy-ins or buy-outs, and allow the scheme to benefit from any upside of the remaining risks present in the scheme. In the same manner as a buy-in, purchasing a longevity swap is an investment decision for the scheme and so is usually obtained for only a subset (typically, the retired members) of the scheme. However, the limited insurance provided by a longevity swap may still be attractive for pension scheme trustees, since they often cite uncertainty in the long-term evolution of mortality rates as a major concern for the scheme.

As with all swap arrangements, the parties to a longevity swap agree to exchange the difference between a fixed and floating series of cashflows. In a longevity swap, the payments comprising the fixed leg of the swap are usually calculated with reference to the best estimate of the projected benefit payments from the scheme in respect of the relevant members.[6] These are typically assessed using an agreed, deterministic set of assumptions for individual mortality rates, as well as other assumptions regarding the rate of pension increases, etc. These best estimate cashflows are then increased by a

[4] The future cashflows for deferred pensioners are more uncertain, since they are of longer term and because deferred members retain options regarding their post-retirement benefits. Therefore, deferred benefits are more expensive to insure.

[5] Longevity swaps are also called "longevity reinsurance" if they are structured as an insurance contract.

[6] Technically, longevity swaps should therefore be called "survivor swaps", as in Dowd et al. (2006a), since what is being swapped is the survivorship of an agreed cohort. In principle, swaps could be constructed using other measures of mortality or longevity, such as probabilities of death ("q-swaps" in the same fashion as q-forwards in Coughlan et al. (2007b)) or period life expectancy. In practice, however, the term "longevity swap" has come to refer uniquely to swaps on survivorship and this usage is adopted in this study.

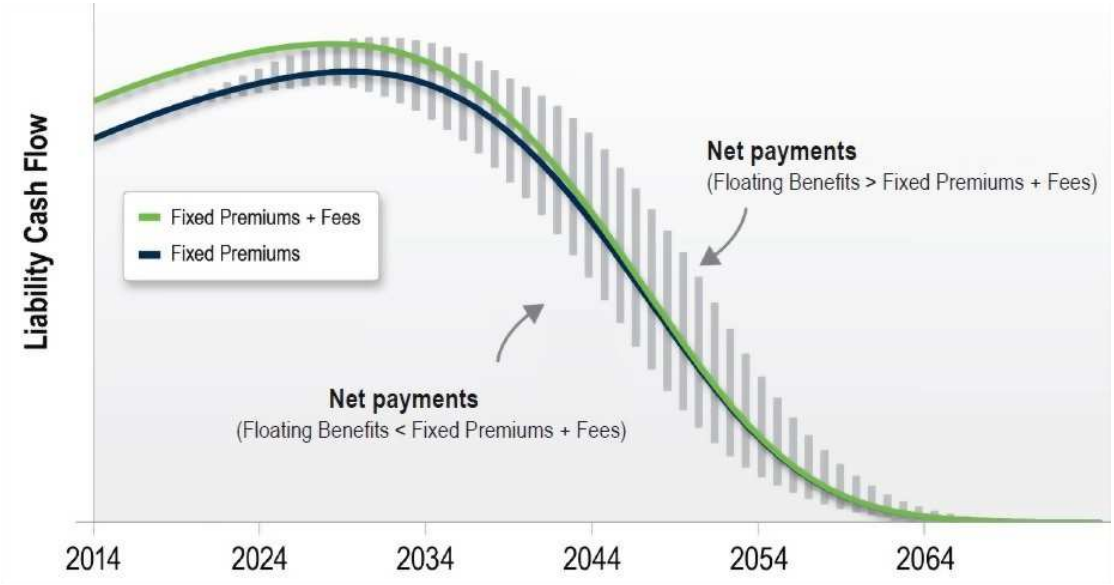


FIGURE 10.1: Illustrative cashflows from a longevity swap (Source: adapted from Kessler (2014))

"longevity swap premium". This premium is set to reflect the degree of risk aversion of both parties, and, anecdotally, swap premiums of between 3% and 5% are not atypical. The floating leg of the swap is set to be equal to the actual benefits paid by the scheme. Such an approach is said to be "bespoke", i.e., tailored to the specific characteristics of the pension scheme.

Figure 10.1, adapted from Kessler (2014), shows an illustrative longevity swap.[7] However, it is important to realise that, in practice, the design of a longevity swap will also need to allow for:

- survivor benefits for potential spouses and dependants of scheme members;
- the different tranches of pension accrued by members (especially in the UK, where different portions of the benefit are subject to different rules for inflationary increases in payment);
- a method for adjusting the fixed leg cashflows to reflect the difference between actual pension increases granted and those assumed at the inception of the swap;[8] and

[7] In Figure 10.1, "fees" refers to the longevity swap premium.

[8] At inception, the fixed leg of the swap will be specified on the basis of a set of assumptions for future increases in pensions in payment. Therefore, the fixed leg will need to be revised subsequently to reflect the differences between this assumption and the actual increases granted by the scheme, in order to ensure that inflation risk is not also transferred in the swap.

- a collateralisation mechanism to reduce the risk of default for either party (see Biffis et al. (2014));

along with various other practical issues. Therefore, longevity swap deals are usually preceded by a lengthy process of negotiation and data cleansing, allowing these practical issues to be resolved before the contract is signed.

Longevity swaps were first developed to transfer risk between life insurers and the capital markets, with the first swap between Friends Provident and Swiss Re in the UK in 2007. Since then, the market has evolved to become dominated by the transfer of risk from pension schemes to life insurance companies and reinsurers, the first being the Babcock-/Credit Suisse transaction in 2009.

In contrast to the bespoke swaps discussed above, much of the academic literature has centred around so called "standardised" or "index-based" longevity swaps. These have the fixed and floating legs of the swap agreement defined with reference to an agreed standard cohort, usually based on the national population. In an index-based swap, the floating leg can be seen as being equivalent to the cashflows from a classic survivor bond (Blake and Burrows (2001)), such as the proposed EIB/BNP Paribas longevity bond in 2004 which has been discussed in previous studies (e.g., Cairns et al. (2006a), Blake et al. (2006) and Lin and Cox (2008)). See Dowd et al. (2006a) and Dawson et al. (2010) for a fuller theoretical discussion of index-based swap agreements.

However, the index-based approach has not been popular with pension schemes to date. The bespoke approach has the advantage that it avoids "basis risk", which arises because systematic differences between the mortality experience of the scheme and of the reference cohort can lead to incomplete risk transfer. Although some studies indicate that basis risk may not be a significant problem (e.g., Coughlan et al. (2011) and Cairns et al. (2013)), there is still a widespread perception that basis risk is an important source of risk in any index-based transaction. Because most pension scheme trustees are highly risk averse, they prefer a more complete risk transfer solution and favour bespoke arrangements.

To date, longevity swaps are mainly targeted at larger pension schemes. Smaller schemes find it more cost effective to conduct a full buy-out, which transfers all the risks associated with running the scheme to a life insurance company, or a buy-in which reduces risk partially. Large schemes, however, may find it difficult to undergo a full risk transfer

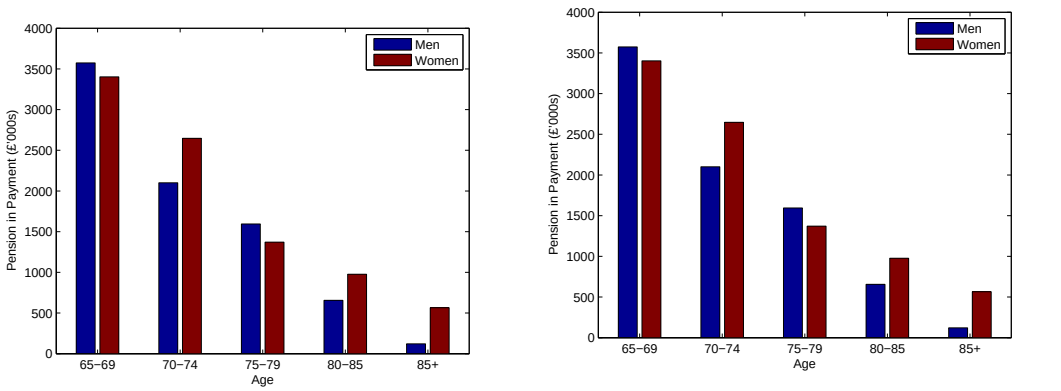
due to capacity constraints within the life insurance sector. Larger schemes also have the ability to take risks, such as investment risk, which they may be rewarded for, and so wish to manage internally.

As counterparties, the majority of the longevity swaps to date have involved either specialised life insurance companies, the insurance subsidiaries of investment banks (who reinsure most of the transferred longevity risk) or directly with reinsurers. Only a relatively small amount of the longevity risk transferred to date has been transferred to the capital markets. It could be argued that longevity risk has become more concentrated and less well diversified in the economy, since it has moved from the balance sheets of dozens of companies and onto the balance sheets of a small number of life insurers and reinsurers. This concentration of risk may, therefore, have increased the risk to macro-economic stability. However, the counter argument to this is that the longevity risk in an occupational pension scheme is held by the corporate sponsor of the scheme, which may not fully understand the nature of the risk. Transferring longevity risk, in contrast, means that it has moved to the strongly regulated and highly capitalised balance sheets of insurance companies which have considerable expertise in managing such risks, and so reduces the threat to macro-economic stability. At the current, early stage of development in the market for longevity risk, it is unclear which of these two considerations is most important.

10.3 The stylised pension scheme

There have been a number of academic studies of longevity risk in pension schemes, for instance, Cossette et al. (2007) and Richards et al. (2013). However, these have analysed far larger schemes than are typical in the UK. The techniques and solutions which are appropriate for such schemes are therefore not directly applicable to the situation in which most UK pension schemes find themselves.

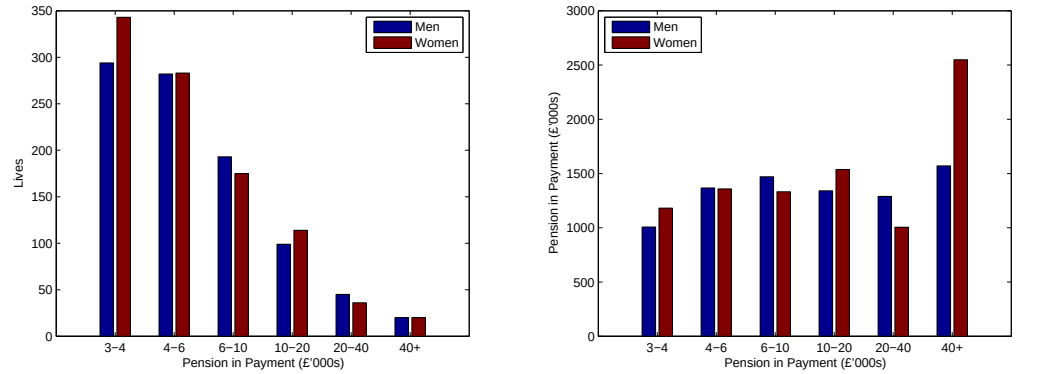
This study considers longevity risk management for a pension scheme more typical of those found in the UK. Membership data for pension schemes is not publicly available and so, for the purposes of this study, we have generated representative member data for a stylised pension scheme. The procedure for doing this has been chosen to reproduce many of the key features of real pension scheme data that are likely to have a significant impact upon longevity risk, as discussed in Appendix 10.A. The advantages of this approach are that we can simplify certain aspects of the complicated benefit structures



(A) Membership numbers by age band

(B) Total amount of pension-in-payment by age band

FIGURE 10.2: Scheme membership by age



(A) Membership numbers by individual pension amount

(B) Total amount of pension-in-payment by individual pension amount

FIGURE 10.3: Scheme membership by individual pension amount

seen in most UK schemes when these are unlikely to be relevant for the management of longevity risk, and we can also avoid problems with data errors and anomalies. The generation of stylised data also overcomes data protection issues, which limit the ability to share and analyse genuine member data.

In this study, we generate a stylised pension scheme with 2,000 members. Figures 10.2 and 10.3 give a summary of the numbers and total pension in payment of men and women in different age and salary bands. In our experience, the patterns shown are typical of UK schemes, with relatively high inequality in the amount of pension in payment caused by the final salary structure.[9] The stylised scheme assumes:

[9] One rule of thumb used in practice, which we have adopted, is that 10% or the members receive 50% of the pension in payment or, equivalently, that 1% of the members receive 25% of the pension in payment. In this regard, UK pension schemes have an income inequality roughly equal to that of the wider UK economy (Institute for Fiscal Studies (2014)).

- equal numbers of men and women retiring at age 65;
- equal average pensions in payment for men and women;
- inflationary increases to pensions in payment;[10]
- no dependants' benefits to spouses and children on the death of members.

We do not believe that any of these assumptions significantly affect out conclusions.[11]

It is important to note, however, that our stylised scheme is still large by UK standards. With 2,000 pensioner members, it would comfortably be amongst the largest 20% of UK pension schemes[12] even without any non-pensioner members. It is, therefore, of a size where longevity risk management solutions, such as longevity swaps are feasible, as discussed in Section 10.2, albeit at the lower end of the range seen to date. This makes it of greater practical interest for modelling compared to smaller schemes which have fewer options to manage their longevity risks.

10.4 Modelling approach

In order to model the longevity and mortality risks in the stylised pension scheme, we start from a set of deterministic "baseline" assumptions for mortality, representative of the assumptions used by pension schemes in the UK for funding or accounting purposes. These assumptions are typically based on standard tables and projections of mortality rates and often do not make any scheme-specific or individual-specific assumptions about mortality rates.

We then move from this set of baseline assumptions to our deterministic "best estimate" assumptions. This set of assumptions consists of a number of different parts. First, we use the "general procedure" of Chapter 5 to construct models of mortality for the national UK population to act as a reference. Then, we use the "relative model" approach described in Chapter 9 to model current mortality rates in the stylised scheme, assuming that they are consistent with those observed in the Self-Administered Pension Scheme

[10] To avoid needing to model inflation, all cashflows shown in this study are expressed in real terms, and a real discount rate is used to calculate present values.

[11] Not allowing for dependants' pensions may understate the impact of longevity risk, since it reduces the term of the liabilities. However, this is offset by assuming equal numbers and equal benefits for men and women: in reality, there are likely to be fewer women than men in a typical pension scheme, who typically receive smaller pensions in payment. Allowing for this would reduce the term of the liabilities relative to what we assume and, hence, these factors will tend to offset each other.

[12] See The Pensions Regulator (2013b).

(SAPS) data. In addition, we then make an allowance for individual mortality rates to vary according to the income of the member. Thus, we can incorporate the features of our stylised scheme, discussed above, into the projected cashflows.

To model the mortality and longevity risks in our stylised pension scheme, we need to go beyond our deterministic best estimate set of assumptions and incorporate uncertainty stochastically. So that our results are internally consistent, we need to ensure that the best estimate assumptions represent the median output of fully stochastic models for each of the component mortality and longevity risks. Consequently, the stochastic models give an equal probability of positive mortality shocks as negative shocks relative to this best estimate. This allows us to separate out our analysis in Sections 10.5 and 10.6 into two parts: the impact of changing the model used to project the most likely scheme cashflows and the riskiness of these cashflows.[13]

For some of the component mortality and longevity risks, we have well-established stochastic models to allow for the uncertainty in projected mortality rates. For instance, systematic longevity risk can be allowed for by projecting the parameters of the reference models for the national population stochastically. However, for other assumptions, such as trend basis risk or the income-related scaling factors applied to specific individuals, no widely-used model exists. To assess the potential impact of uncertainty in these assumptions, we make more approximate allowances, in line with our own practical experience of buy-out, buy-in and longevity swap transactions. Whilst the specific details of these allowances may appear ad hoc, we have taken steps to ensure that their impact on the projected cashflows is broadly reasonable. However, we believe that further research into these subjects is necessary.

10.4.1 The baseline set of assumptions

As a set of baseline assumptions we use:

- Male and female mortality rates in 2008 given by the S2PMA and S2PFA mortality tables, graduated in Continuous Mortality Investigation (2014a) from data weighted on an "amounts" basis (see Section 10.4.2.3) from the SAPS study. These tables are typical of those used by pension schemes in the UK for accounting and

[13] We feel that this distinction is often overlooked, as "longevity risk" is sometimes used to describe the impact of moving from inappropriate deterministic assumptions to more realistic ones (e.g., Antolin (2007) and Oppers et al. (2012)), as opposed to the uncertainty in the realistic assumptions.

funding purposes (The Pensions Regulator (2013a) and Sithole et al. (2012)), but do not allow for scheme-specific or individual-specific mortality effects.

- Improvements in mortality rates are given by the CMI Projection Model[14] with a long-term rate of improvement of 1.5%. This model is widely used in the UK and has become the benchmark method of projecting mortality for funding and accounting purposes (for instance, see The Pensions Regulator (2013b)) and the long-term rate of improvement is broadly consistent with the assumption used for funding and risk assessment purposes.

10.4.2 Modelling mortality and longevity risks

We classify the component mortality and longevity risks present in the stylised pension scheme into three broad categories, with separate (but inter-related) modelling approaches for each.

1. First, in Section 10.4.2.1, we consider mortality rates in the national population in order to model the systematic components of longevity risk. In order to do so, we use "reference models" constructed using the "general procedure" described in Chapter 5.
2. Second, in Section 10.4.2.2, we investigate the scheme-specific longevity basis risks present, in order to consider the ways in which mortality rates may be different in the pension scheme compared to the national population. We do this via a "relative" modelling approach, as discussed in Chapter 9.
3. Finally, in Section 10.4.2.3 we allow for individual-specific mortality risks, such as income-related scaling factors adjusting the mortality rates for an individual scheme member and the idiosyncratic mortality risk in the timings of individual deaths.

Each of these components can modelled stochastically to assess the magnitude of the mortality and longevity risks present in the scheme, or deterministically to obtain the best estimate set of assumptions described above.

10.4.2.1 The reference models for the national population

To model mortality rates in the UK national population, we use data from Human Mortality Database (2014) and "reference models" constructed using the "general procedure"

[14] Described in Continuous Mortality Investigation (2009a,b) and updated in Continuous Mortality Investigation (2013).

of Chapter 5 for each sex. These models are of the form

$$\ln\left(\mu_{x,t}^{(R)}\right) = \alpha_x^{(R)} + \sum_{i=1}^{N} f^{(R,i)}(x;\theta^{(R,i)})\kappa_t^{(R,i)} + \gamma_{t-x}^{(R)} \tag{10.1}$$

where

- $R \in \{UKm, UKf\}$, i.e., we fit separate models for male and female mortality data.
- age, x, is in the range $[50, 100]$, period, t, is in the range $[1950, 2011]$ and, therefore, that year of birth, y, is in the range $[1850, 2010]$;
- $\alpha_x^{(R)}$ is a static function of age;
- $\kappa_t^{(R,i)}$ are period functions governing the evolution of mortality with time;
- $f^{(R,i)}(x;\theta^{(R,i)})$ are parametric age functions (in the sense of having a specific functional form selected a priori) modulating the impact of the period function dynamics over the age range, potentially with free parameters $\theta^{(R,i)}$;[15] and
- $\gamma_y^{(R)}$ is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as it ages.

The general procedure selects the number of age/period terms, N, and the form of the age functions, $f^{(R,i)}(x)$, in order to construct mortality models which give a close but parsimonious fit to the data. This way, we aim to extract as much information as possible from the national population dataset and have specific terms within the model corresponding to the different features of interest. This procedure was performed on male and female mortality data from the UK for ages 50 to 100 in Chapter 9, where a full description of the final models and the tests performed on them can be found. In that study, the general procedure selected models with three age/period terms for both men and women of the forms given in Table 10.1.[16]

Systematic longevity risk

[15] For simplicity, the dependence of the age functions on $\theta^{(R,i)}$ is supressed in notation used in this study, although it has been allowed for when fitting the model to data.

[16] Demographic significance, as used in Table 10.1, is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

Term	Men		Women	
	Description	Demographic Significance	Description	Demographic Significance
$f^{(R,1)}(x)\kappa_t^{(R,1)}$	Constant age function	General level of mortality	Constant age function	General level of mortality
$f^{(R,2)}(x)\kappa_t^{(R,2)}$	"Call" age function	Older age mortality	"Call" age function	Old age mortality
$f^{(R,3)}(x)\kappa_t^{(R,3)}$	"Put" age function	Younger age mortality	Gaussian age function	Younger age mortality

TABLE 10.1: Terms in the reference models constructed using the general procedure for UK men and women ages 50 to 100

To project mortality rates in the national population, we use a random walk with drift for the different period functions

$$\begin{aligned}\boldsymbol{\kappa}_t &= \left(\kappa_t^{(UKm,1)}, \quad \ldots \quad \kappa_t^{(UKm,3)}, \quad \kappa_t^{(UKf,1)}, \quad \ldots \quad \kappa_t^{(UKf,3)}\right)^\top \\ \boldsymbol{\kappa}_t &= \boldsymbol{\kappa}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\epsilon}_t \end{aligned} \tag{10.2}$$

and an AR(1) around linear drift process for the cohort parameters

$$\begin{aligned}\boldsymbol{\gamma}_y &= \left(\gamma_y^{(UKm)}, \quad \gamma_y^{(UKf)}\right)^\top \\ \boldsymbol{\gamma}_y - \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1 y &= R(\boldsymbol{\gamma}_{y-1} - \boldsymbol{\beta}_0 - \boldsymbol{\beta}_1(y-1)) + \boldsymbol{\varepsilon}_y \end{aligned} \tag{10.3}$$

By using multivariate time series of this form, we allow for any correlation in mortality improvements between men and women in the UK which is observed in the historical data. These time series processes have been chosen to be "well-identified" in the sense of Chapters 3 and 4, i.e., the projected mortality rates are independent of the identifiability constraints imposed upon the model.

To give deterministic best estimate assumptions, we set $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\varepsilon}_y$ to be equal to zero in future. We refer to the variation generated by allowing $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\varepsilon}_y$ to be (normally-distributed) random variables we refer to as "systematic longevity risk", because it affects all members of the UK national population and so cannot be reduced by pooling or diversification.

Parameter uncertainty

In addition to systematic longevity risk, we also investigate the impact of parameter uncertainty in the reference population. This is the uncertainty due to the fact that the parameters in the reference model and the time series processes are not known with certainty, but are estimates based on finite data. We do not anticipate parameter uncertainty in the reference population to be particularly large, since there is a lot of data available for the national UK population. However, it is important to allow for parameter uncertainty to avoid hierarchical issues in the relative model (as discussed in Chapter 9), which are due to the parameters in the relative model being estimated conditional on the previously estimated parameters of the reference population. To allow for parameter uncertainty, we use the residual bootstrapping procedure of Koissi et al. (2006) to generate multiple realisations of the parameters in the reference model, which are then used to re-estimate the parameters of the time series process in Equations 10.2 and 10.3.

10.4.2.2 The relative models for the scheme

The next stage of the modelling process is to investigate the scheme-specific factors which can influence mortality rates. We call this the "basis" for the scheme. We decompose this basis into two parts:

1. the differences in the current level of mortality rates between the national population and the scheme, which we call the "level basis"; and
2. the differences in the rates of change in mortality rates between the national population and the scheme, which we call the "trend basis".

The uncertainty in the measurement of these two parts we refer to as the "level basis risk" and "trend basis risk", respectively.

Before we begin to model the basis for the stylised pension scheme, we must first simulate exposures to risk and death counts for the scheme, which we can then fit a model to. We assume that the stylised scheme is typical of the SAPS population and, therefore, use data from the SAPS study in order to estimate the parameters in the relative model.[17] However, since the SAPS dataset is far larger than any occupational UK pension scheme, we need to rescale this data to make it comparable with the size of the stylised scheme. To do this, we assume that the scheme membership has remained constant at each age for a period of twelve years prior to 2011 (i.e., the period of the SAPS data) to give

[17] We are indebted to the Continuous Mortality Investigation for providing this data. For further details see Chapter 9.

exposures to risk.

We then generate best estimate death counts for the scheme using these exposures using the observed mortality rates in the SAPS data, i.e.,

$$D_{x,t}^{(S)} = E_{x,t}^{(S)} m_{x,t}^{(SAPS)}$$

where $E_{x,t}^{(S)}$ are the assumed central exposures to risk at each age and year, and $m_{x,t}^{(SAPS)} = \frac{D_{x,t}^{(SAPS)}}{E_{x,t}^{(SAPS)}}$ are the central mortality rates observed in the SAPS populations.

This procedure gives us simulated data for the stylised scheme that is consistent with that from the SAPS study, which we can then use to model the basis for the stylised scheme. To do this, we use the "relative" approach developed in Chapter 9. This proposes a model of the form

$$\ln\left(\mu_{x,t}^{(S)}\right) = \alpha^{(R)} + \alpha_x^{(\Delta)} + \sum_{i=1}^{N} \lambda^{(i)} f^{(R,i)}(x)\kappa_t^{(R,i)} + \lambda^{(\gamma)}\gamma_{t-x}^{(R)} + \nu X_{t-x} \tag{10.4}$$

where $\alpha_x^{(\Delta)}$ is the difference in the level of mortality between the two populations[18] and the $\lambda^{(j)}$ ($j \in \{1, 2, 3, \gamma\}$) correspond to the "sensitivity" of the small population to the factor j in the reference population.[19] Therefore, the definitions above imply that $\alpha_x^{(\Delta)}$ controls the level basis for the scheme, whilst the $\lambda^{(j)}$ control the trend basis. Level basis risk and trend basis risk correspond to the uncertainty in estimating these parameters.

Level basis

Based on the results of Chapter 9, we restrict $\alpha_x^{(\Delta)}$ to be of parametric form, i.e.,

$$\alpha_x^{(\Delta)} = \sum_{i=1}^{N+1} \alpha^{(i)} \tilde{f}^{(R,i)}(x)$$

where $\tilde{f}^{(i)}(x)$ is an expanded set of the age functions present in the reference model plus an additional linear function required for identifiability.[20] This means that, instead of estimating separate values of $\alpha_x^{(\Delta)}$ at each age, there are only $N+1$ components, $\alpha^{(i)}$,

[18] For example, mortality rates in the scheme at age x might be consistently 5% lower than those in the reference population for all times.

[19] For example, the scheme may experience 90% of the change due to $\kappa_t^{(R,1)}$ in the national population.

[20] See Appendix 9.B of Chapter 9 for a discussion of why this is necessary.

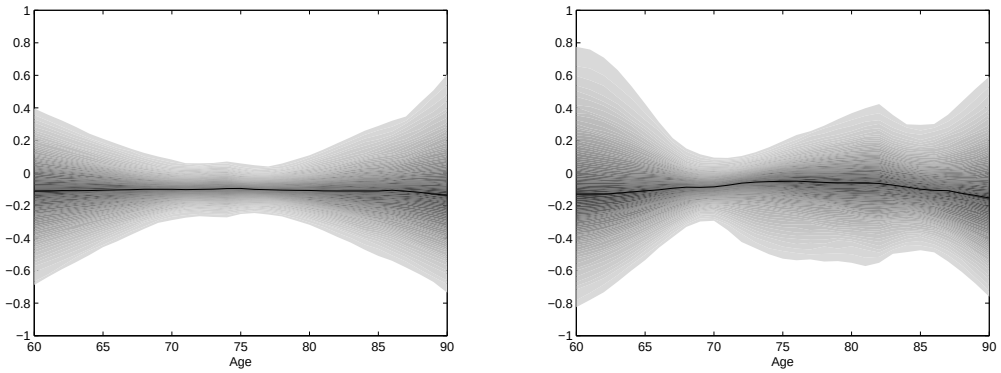
for the level basis for the scheme for each sex, which makes the model considerably more parsimonious. In Chapter 9, this was found to be necessary to avoid over-parameterising the model, especially for small population sizes.

This approach is conceptually similar to the standard actuarial practice of specifying a base mortality table by making a series of adjustments (given by $\alpha_x^{(\Delta)}$) to a standard mortality table (in this case, given by the reference models for the national population). The estimation of $\alpha_x^{(\Delta)}$ in this study is conducted on a purely statistical basis, comparable to performing an analysis of the experience data of the scheme. In practice, however, specifying the base table will also make use of more subjective adjustments, e.g., to reflect the industry scheme members were employed in. Furthermore, the involvement of a life insurer, with access to greater volumes of data and more sophisticated modelling techniques, is likely to reduce the uncertainty in specifying the base table (i.e., the level basis risk) considerably from what could be achieved by the scheme alone.

To allow for level basis risk, we adopt a similar approach to that used to allow for parameter uncertainty in the reference population, i.e., we use a residual bootstrapping procedure based on Koissi et al. (2006). To do this, we take the residuals from fitting the relative model to the data for the scheme and use these to generate random death counts for the stylised scheme. To these, we refit the relative model in Equation 10.4 to generate new estimates of $\alpha_x^{(\Delta)}$.[21] Figure 10.4 shows the 95% confidence intervals for $\alpha_x^{(\Delta)}$ found using this procedure.

Since $\alpha_x^{(\Delta)}$ is restricted to be a linear combination of age functions, the pattern of level basis risk across ages depends strongly upon the form of those age functions. However, we can see that the uncertainty is greatest at the highest and lowest ages in the range, due to the very low absolute numbers of deaths expected at these ages. We also see that, at most ages, $\alpha_x^{(\Delta)}$ is unlikely to be more than ± 0.1 from its best estimate value. This corresponds to a relative uncertainty in the level of mortality rates of around 10% at any age. However, there is a considerable "tail" to this distribution, which means that we are unable to rule out significantly higher or lower levels of mortality rates in the scheme compared with the national population. Comparing Figure 10.4 with Figure 9.5 in Chapter 9, we note that the basis risk for the scheme is significantly greater than for the full SAPS population, due to its relatively small size. It is interesting to note that,

[21] In addition, we use a bias correction technique to ensure consistency between this procedure and the deterministic best estimate assumption, since otherwise the lower bound of zero deaths at any age can give anomalous results.



(A) Men

(B) Women

FIGURE 10.4: 95% fan charts showing the parameter uncertainty in $\alpha_x^{(\Delta)}$ (level basis risk)

while experience studies conducted to quantify the level basis in pension schemes are becoming more common, it is less usual to see the uncertainty in the level basis quantified. These results indicate that this uncertainty in the level basis may be substantial, even for a comparatively large pension scheme.

It is also interesting to consider the technique proposed in the Solvency II standard model for systematic longevity risk (EIOPA (2014)), which is to reduce the level of mortality in the scheme by 20%. A common criticism (e.g., Nielsen (2010) and Börger (2010)) of this approach is that it is a poor proxy for systematic longevity risk, which is likely to emerge slowly over time rather than immediately as a one-off shock. However, the Solvency II standard model for systematic longevity risk could be considered as a scenario for investigating the impact of level basis risk. We see from Figure 10.4 that a 20% reduction in mortality rates lies within the 95% confidence intervals for level basis risk for most ages. Therefore, the model proposed by EIOPA (2014) can be considered a reasonable proxy for investigating level basis risk in a pension scheme, despite its shortcomings as a proxy for systematic longevity risk.

Trend basis

We now consider the potential trend basis in the stylised pension scheme. However, doing so is very difficult because quantifying trend basis requires far more data than any pension scheme is likely to have, as discussed in Chapter 9. Attempting to do so using the relative model in Equation 10.4 would lead to an over parameterised model, and parameter estimates which are not robust (i.e., have very large parameter uncertainty). This would then lead to unfeasibly large estimates of the trend basis risk, which can be

ruled out on the grounds of biological reasonableness.[22] For example, experiments with this approach have led to scenarios where life expectancy at age 65 in the stylised scheme rises rapidly to over 40 years or drops precipitously to almost zero.

For small populations of the same size as our stylised scheme, the relative modelling approach in Chapter 9 showed a very strong preference for restricting the model so that $\lambda^{(j)} = 1$ for each of the age/period and cohort terms for both men and women. This made the model considerably more parsimonious and robust when fitting it to data. For the purposes of this study, we impose the same restriction, which is equivalent to assuming that there is no trend basis in the scheme. Because this makes the model more robust, it also reduces the uncertainty in the estimation of $\alpha_x^{(\Delta)}$ (i.e., the level basis risk) compared with using an over-parameterised model.

Imposing $\lambda^{(j)} = 1$ is equivalent to imposing a priori that there is no trend basis and no trend basis risk between the reference and sub-populations. However necessary this assumption is when obtaining the best estimate set of assumptions, we will need to relax it and allow $\lambda^{(j)}$ to vary when performing stochastic projections to estimate the trend basis risk. To do this, we use an informal procedure based on our desire for biological reasonableness. Various studies, such as Lu et al. (2012) (especially Tables 3 and 4) and Haberman et al. (2014), have indicated that the magnitude of differences between the trend rate of improvement in mortality rates between various sub-populations and the national population is of the order of 0.5% p.a..[23]

To generate trend basis risk of around the correct magnitude, we allow $\lambda^{(1)}$ to vary using

$$\lambda^{(1)} \sim N(1, \sigma_\lambda^2)$$

where $\sigma_\lambda \approx 0.3$. This also imposes $\mathbb{E}\lambda^{(1)} = 1$, to ensure that the results of this procedure are consistent with our deterministic best estimate. Although this procedure is somewhat informal, we are confident that we obtain results which are biologically reasonable and are consistent with the findings in the studies mentioned above. For simplicity, the other scaling factors in the relative models are not allowed to vary stochastically and so

[22] The concept of biological reasonableness was introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*"

[23] In the reference models used in this study, the expected rates of improvement are around 1.5% p.a. in the reference population for both sexes, given by the drift of $\kappa_t^{(R,1)}$ from the random walk process in Equation 10.2.

are set equal to unity.

10.4.2.3 Individual mortality risks

Going beyond the national and scheme-specific evolution of mortality, we also need to consider individual-specific features of mortality rates, such as the specific mortality rates for each individual scheme member and their random time of death, which we refer to as mortality risks in the stylised scheme.

Individual income-related scaling factors

Mortality rates are likely to be different for different individuals, since wealth and lifestyle factors have an impact on longevity. Some of these factors will already be taken into account at the scheme level through the analysis of the basis. However, a pension scheme is not a homogenous group of individuals, and this may have important consequences in any assessment of the risks faced by the scheme. Of these factors, the correlation between income and life expectancy will probably be the most important in modelling the stylised scheme, since individuals who are in receipt of the largest pensions contribute most to the total scheme cashflow.[24]

In practice, these factors are often taken into account by using mortality rates from standard tables which have been estimated on an "amounts" basis. This approach weights the experience of each life under observation by the amount of pension in payment, and so will give more weight to the highest income pensioners. This means that tables estimated on an amounts basis tend to give lower mortality rates than tables estimated on the same data on a "lives" basis, i.e., where all lives under observation are given equal weight. Such an approach will give mortality rates that are appropriate for evaluating liabilities on an aggregate basis (because the weight each life receives in the liabilities is also proportional to their pension amount). However, mortality rates estimated on an amounts basis will not be appropriate for any specific individual, which may bias the results of any member-by-member risk assessment for the stylised scheme.

[24] We implicitly assume that an individual's pension is their only source of income and, therefore, that income and pension amount are synonymous.

Alternatively, the correlation between income and longevity can be allowed for on an individual basis by using mortality tables estimated on a lives basis (such as given by the relative modelling approach in Section 10.4.2.2) and then using individual scaling factors, i.e., introducing factors, K_j for each individual j, which scale the mortality rates experienced by an individual relative to the average scheme mortality

$$\mu_{x,t,j} = K_j \mu^{(S)}_{x,t} \tag{10.5}$$

It is important to note that these scaling factors are relative to the aggregate scheme mortality rates, which are, themselves, unknown. In some respects, these can be considered as analogous to the "frailty" factors in Vaupel et al. (1979) or the results of performing a Cox proportional hazard model (Cox (1972)).

Since we are interested in allowing for the individual mortality risks in our stylised pension scheme as well as the systematic and scheme specific risks, we adopt the latter approach and use individual scaling factors to allow for the correlation between income and longevity. In practice, the individual scaling factors are often found by conducting a "postcode analysis", where information on the address of the individual is used to make inferences about their wealth and lifestyle.

However, because our stylised scheme is purely illustrative, we are not able to perform an actual postcode analysis. Instead, we use an approximate set of scaling factors, which are broadly consistent with the magnitudes of the income-related scaling factors in Villegas and Haberman (2014) and Continuous Mortality Investigation (2012), and are consistent with our practical experience of the results of actual postcode analyses. These scaling factors are based solely on income, with an assumption that:

- The quintile of scheme members receiving the largest pensions at retirement (pensions over over £9,250 p.a. in our modelling) experience mortality rates 70% of the average for the scheme;
- The quintile receiving the lowest pension amounts (between £3,000 and £3,500 p.a. in our modelling) experience mortality rates 130% of average; and
- The individual scaling factor are linearly interpolated between 70% and 130% for the middle quintiles.

This means that the median income level (£4,900 p.a.) corresponds to an individual scaling factor of 100%. This has the consequence that the individual scaling factors do

not systematically bias the mortality rates that would be observed in the scheme when all members are given equal weight, i.e., those given by an analysis conducted on a lives basis. These individual scaling factors are shown in Figure 10.5.

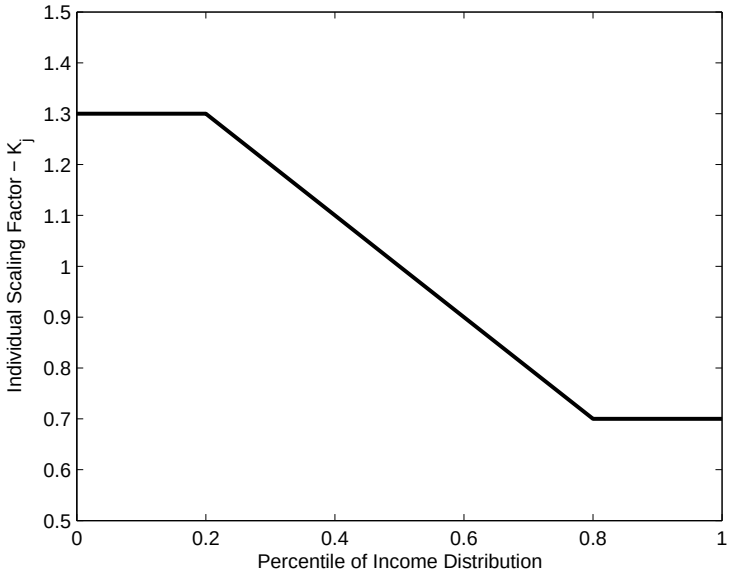


FIGURE 10.5: Individual income-related scaling factors

To allow for uncertainty in the individual scaling factors in the stochastic model, we assume

$$K_j = K_{j,\text{Best Estimate}} \times \exp(Z_j)$$
$$\text{where} \quad Z_j \sim N\left(0, 10\%^2\right)$$

Since the scaling factors are multiplicative, this assumption avoids the possibility that individual mortality rates could be negative. It also ensures that the median of our stochastic simulations corresponds to the best estimate scaling factors discussed above. However, it is important to note that, just as with the best estimate of these income-related scaling factors, the risk attached to them is illustrative. Even when a postcode analysis is performed on genuine member data, the uncertainty in the scaling factors is rarely (if ever) quantified. However, we believe that our approach is reasonable, since an error of 0.1 on the scaling factor for an individual is comparable to an error of ten percentage points of the income distribution to which the member belongs.[25] This is realistic given the multiple sources of income that pensions scheme receive in practice.

[25] E.g., a scaling factor of 110% as opposed to 100% would place a member in the 40^{th} percentile of the income distribution, rather than the 50^{th} percentile.

Idiosyncratic risk

The models above allow us to estimate the mortality rates experienced by a scheme member, by looking first at the national population, then at the scheme itself and finally on an individual basis, along with estimating the uncertainty in these estimates arising at each stage. However, even if the mortality rates were known with certainty, the time of death of any specific individual (and hence the total benefits paid to them) would still be uncertain. We refer to this uncertainty as "idiosyncratic risk".

In principal, idiosyncratic risk can be diversified away and so should not be a significant risk for a suitably large scheme - see Milevsky et al. (2006). The stylised scheme in this study has 2,000 members, which is large by the standards of UK pension schemes. However, it is still important to allow for idiosyncratic risk since the stylised scheme contains a minority of members with large pensions, for whom the exact time of death will still have an important impact on the projected benefits paid by the scheme.

We allow for idiosyncratic risk by considering each member individually, with the random future lifetime modelled as an inhomogeneous-Poisson process subject to a hazard rate given by their modelled mortality rates.

10.5 Establishing the best estimate of scheme cashflows

As discussed at the start of Section 10.4, the first stage in our modelling approach is to move from the baseline set of assumptions, typical of those used by pension schemes in the UK for funding purposes, to the best estimate assumptions found from the model. Quantifying the impact of changing these assumptions is useful in assessing the potential for misspecification of the fixed leg of the swap. Since the best estimate assumptions are those which give an equal probability of positive and negative mortality shocks impacting the future scheme cashflows (and hence the floating leg), potential misspecification can lead to systematic bias the net cashflows from the swap in favour of either the pension scheme or the swap provider.

To quantify the impact of this potential bias, we change each of the baseline assumptions in turn and independently, i.e.,

1. We use the mortality rates fitted by the relative model for the scheme in 2008 instead of those given by the S1PMA and S1PFA mortality tables.

2. We use the best estimate projections of mortality for the national population with no trend basis to project mortality rates, instead of the CMI projection model.

3. We use the income-related scaling factors to adjust individual mortality rates, rather than using the scheme mortality rates for all members.

Figure 10.6 and Table 10.2 show the impact of these factors on the present value and duration of the scheme cashflows, individually and in aggregate, using a real discount rate of 1.0% p.a..

	PV (£m)	Δ PV	Duration (years)	Δ Duration
Baseline	284.7	-	10.6	-
2008 mortality rates	273.0	-11.7	10.3	-0.3
Projection model	274.5	-10.2	10.2	-0.4
Individual scaling factors	305.5	20.8	11.3	0.7
Best estimate	290.5	5.8	10.9	0.3

TABLE 10.2: Present values and durations of scheme cashflows on the baseline and best estimate sets of assumptions

As can be seen from Table 10.2, the present values of the liabilities are not significantly different under the baseline and best estimate assumptions, with a total difference of only £5.8m, or 2% of the present value of the liabilities. Looking at the pattern of cashflows in Figure 10.6, we see that the biggest differences in the projected cashflows under the baseline and best estimate sets of assumption occur after 30 or so years of projection (i.e., after 2040), and so are heavily discounted and make relatively little difference to the present value. In many ways, this is reassuring as it implies that the deterministic assumptions used by schemes for funding purposes are not substantially overestimating or underestimating the liabilities compared with what could be obtained using more sophisticated models. However, it is interesting to see that this is only true in aggregate, and that the specific mortality assumptions can make sizeable differences to the present value of the liabilities.

First, we see that the CMI Projection Model with a long-term rate of improvement of 1.5% p.a. slightly overstates the projected improvements in mortality compared with the reference model described in Section 10.4.2.1, since it give a present value for the liabilities £10.2 higher than the best estimate assumption. This is because, whilst the best estimate assumption also gives improvements in mortality rates of around 1.5% p.a.,

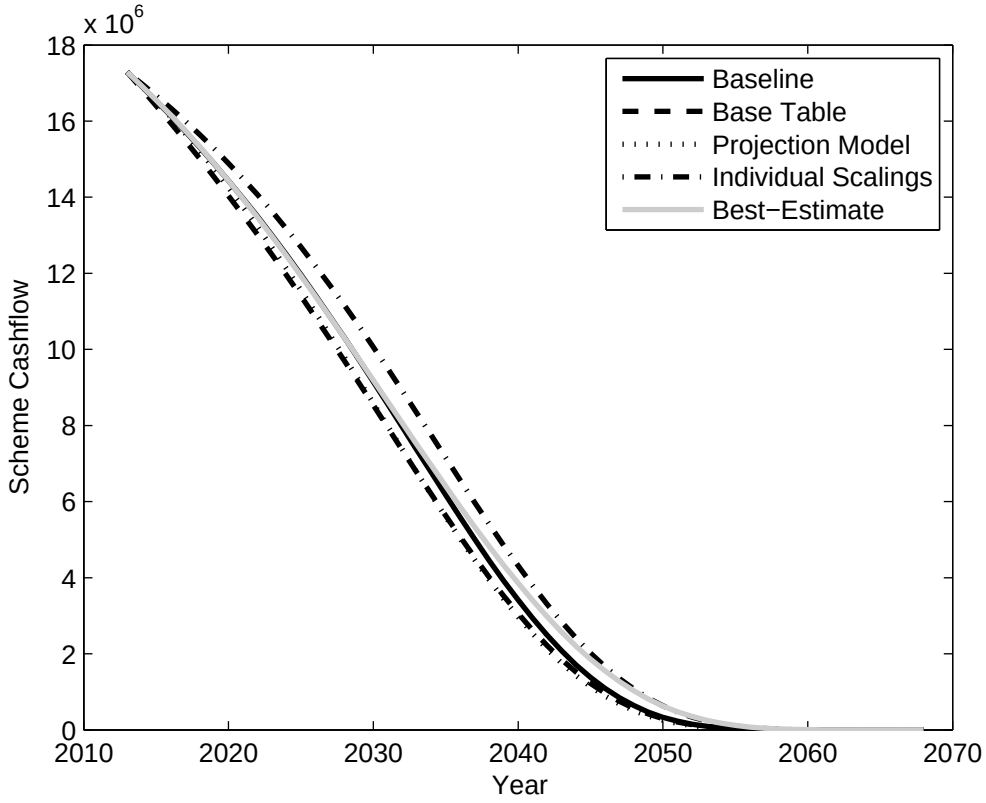


FIGURE 10.6: Projected deterministic cashflows using different sets of assumptions

the pattern of improvements across different ages, cohorts and future years can be very different to that given by the CMI Projection model. However, since the CMI Projection Model uses a fundamentally different approach to project mortality from that used by the reference model, it is reassuring that the difference in the liabilities between the two models is relatively small.

In addition, we see from Table 10.2 that the impact of moving from the baseline assumption for current mortality rates - i.e., moving from using mortality tables graduated from the SAPS data on an amounts basis for all members to mortality rates from the relative model estimated from the SAPS data on a lives basis with individual adjustments to reflect the amount of pension in payment - broadly offset each other. This implies that, in aggregate, the common practice of using standard tables graduated on an amounts basis gives a reasonable estimate of the liabilities compared with one with individual scaling factors. Conversely, it also implies that the relatively crude method of obtaining individual scaling factors used in Section 10.4.2.3 (which was chosen to be consistent with our experience of postcode mortality studies in practice) broadly replicates the observed relationship between income and longevity found in the SAPS data in aggregate.

10.6 Assessing and comparing different sources of risk

We now introduce the different sources of mortality and longevity risk described in Section 10.4 to assess their impact and measure the variability of the cashflows on the best estimate set of assumptions shown in Section 10.5. To quantify this, we estimate the standard deviation of the present value of scheme cashflows (i.e., the deviation around the best estimate value of £290.5 shown in Table 10.2). This gives us a broad measure of the total uncertainty in the future cashflows arising from the different sources.

For a set of pension scheme trustees considering a bespoke longevity swap, another key considerations is the insurance value of the swap. Assuming that the distribution of the projected cashflows is roughly symmetrical at any future time, an actuarially fair swap will have the fixed leg of the swap equal to the best estimate of the future scheme cashflows. This would ensure that there will be an equal probability of the floating leg being greater than or less than the fixed leg (i.e., of a positive or negative net cashflow from the swap) and, hence, the swap would have zero expected present value for both parties. In practice, the fixed leg cashflows are set by increasing the best estimate cashflows by the swap premium (which we have set at 4%, consistent with our practical experience of swap arrangements), which means that the swap has positive expected present value for the provider and negative expected present value for the scheme. This reflects the premium the scheme is willing to pay to transfer risk to the swap provider.

A consequence of this is that, in the short term, there is a high probability that the scheme will make net payments under the swap arrangement (since the short-term cashflows will be the most certain and so unlikely to be in excess of the fixed leg). Therefore, the trustees may find it hard to explain the value of the swap as an insurance policy over the longer term to the other stakeholders of the pension scheme and, hence, have difficulty justifying entering into the swap.

To illustrate the insurance value of the swap, we calculate the probability of the scheme receiving a positive net payment from the swap in year t i.e.,

$$P(t) = \mathbb{P}\left[C_t - 1.04C_t^{\text{Best Estimate}} \geq 0\right] \tag{10.6}$$

where C_t is the projected cashflow from the scheme, for $t \leq 20$. We consider the estimated values of $P(t)$ for only the first twenty years of the swap arrangement because

most of the liabilities are expected to have run off by the end of this period and it approximates the upper limit of the timescales being considered by the decision makers when a swap is being considered.

However, it is important to note that the insurance value of a longevity swap does not solely depend upon the probability of the scheme receiving a positive net payment. The swap also has value even if no positive net payments are made, since it allows the scheme to fix the effective mortality rates experienced for the members covered by the contract. This may be especially desirable for schemes which have had to change their assumptions for future mortality rates at successive funding valuations (almost always causing an increase in the liabilities) and are willing to pay a premium to lock into a specific set of assumptions which will not need revision going forwards and so obtain certainty over mortality rates.

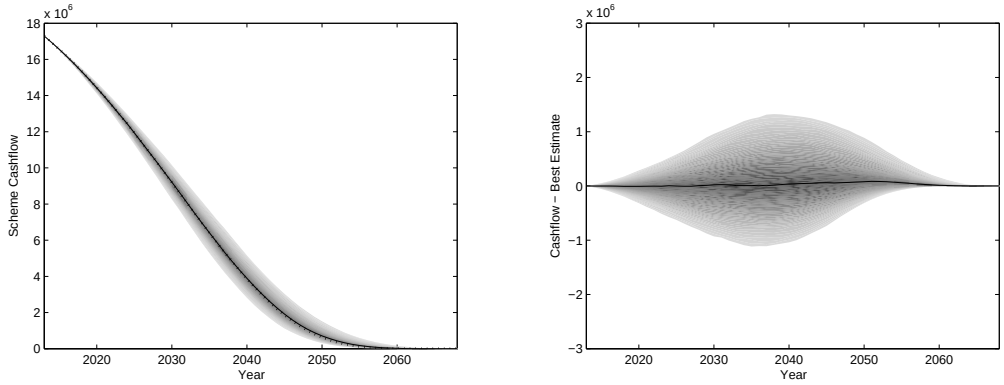
We assess the different mortality and longevity risks in the stylised scheme in two stages. First, each source of risk is considered in isolation, setting all the other sources of risk equal to their best estimates, to assess its relative importance.[26] Once the risk sources have been considered independently, all of the risks are combined to fully assess the potential mortality and longevity risks within the stylised scheme and, hence, the ability of a longevity swap to transfer them effectively. It is important to note that, because the allowance for many of these risks is quite approximate, our results are subject to considerable model risk.[27]

10.6.1 Systematic longevity risk

In Section 10.4, we defined systematic longevity risk as the risk arising from the stochastic projection of the period and cohort functions for the reference UK population using the time series processes in Equations 10.2 and 10.3. Figure 10.7a shows the 95% projection intervals for the projected (real) cashflows of the scheme allowing for this risk, where the median value is equal to the cashflows on the best estimate set of deterministic assumptions shown in Figure 10.6. To highlight the pattern of uncertainty in the projected cashflows, Figure 10.7b shows the difference between these cashflows and the best

[26] For instance, to allow for systematic longevity risk, we project the period and cohort parameters for the reference populations stochastically using Equations 10.2 and 10.3, but do not allow for parameter uncertainty, set the parameters of the relative models equal to their best estimate (without any allowance for uncertainty and, hence, basis risk), use the best estimate individual scaling factors and do not allow for idiosyncratic risk.

[27] Defined as the uncertainty caused by our model being an approximation to the true underlying processes governing the phenomenon in question, as discussed in Cairns (2000).



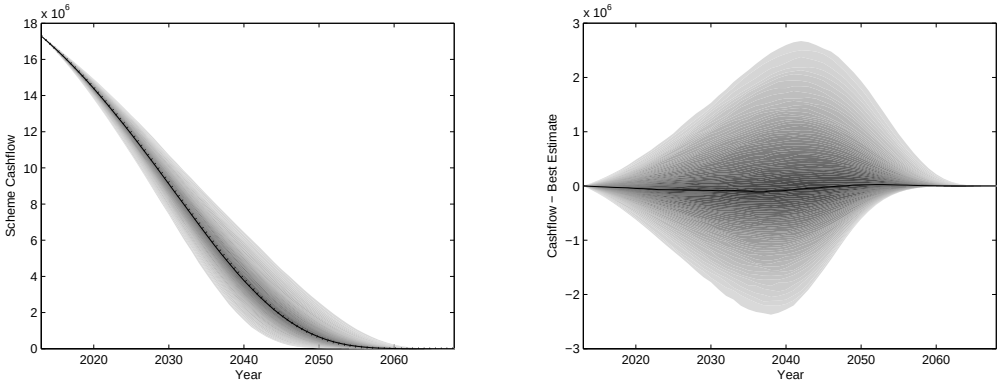
(A) Scheme cashflows (B) Net cashflows

FIGURE 10.7: Impact of systematic longevity risk on projected scheme cashflows

estimate (i.e., the net payments from a swap with no premium). Thus, the magnitude of the uncertainty in the projected cashflows for any future year can be assessed. Overall, the standard deviation of the present value of the scheme cashflows (using a real discount rate of 1.0%) due to systematic longevity risk is £13.6m, or c. 4.7% of the best estimate present value of £290.5m shown in Table 10.2.

We see that the main impact of allowing for stochastic mortality projections is in cashflows due to take place in around twenty years' time. This is not surprising, given that it takes time for the uncertainty in the projected mortality rates due to systematic longevity risk to give a noticeable effect. This is because it first takes time to first generate larger differences in mortality rates, and only then, when these are significant, compound these differences in mortality rates into significant differences in the projected cashflows.

We also find that, looking at systematic longevity risk alone, the probability that the scheme receives a positive net cashflow, $P(t)$ in Equation 10.6, stays relatively low (close to zero) for the first eleven years of the swap arrangement, but then grows steadily beyond this. This may be of interest, since systematic longevity risk is often a major concern to pension scheme trustees, and so they need to be aware that a longevity swap only has significant insurance value against this risk after a decade or so. In addition, an index-based longevity swap (which would only offer protection against systematic longevity risk) would have similar issues and so could only be regarded as providing insurance only over longer time periods.



(A) Scheme cashflows (B) Net cashflows

FIGURE 10.8: Impact of level basis risk on projected scheme cashflows

10.6.2 Parameter uncertainty

Parameter uncertainty in the reference population has only a very small impact on the projected cashflows of the scheme, with a standard deviation in the present value of the cashflows of £0.5m (0.2% of the best estimate present value). This is unsurprising, for the reasons discussed in Section 10.4.2.1, namely that the reference population is large and therefore gives reliable parameter estimates. However, it is important to allow for parameter uncertainty in the reference population due to the hierarchical nature of the relative model, as discussed previously.

10.6.3 Level basis risk

The impact of level basis risk on the projected scheme cashflows is shown in Figure 10.8. The standard deviation of the cashflow present value is £27.4m, or c. 9.4% of the best estimate value, which is large relative to the other risks we investigate. This should not be surprising, given that the high degree of uncertainty shown in Figure 10.4 and the fact the level basis risk will impact the scheme cashflows immediately, rather than taking time to develop as macro-longevity risk and trend basis risk do.

One interesting feature in Figure 10.8b is the asymmetry of the confidence intervals, as shown by the peak of the "downside" risk from the point of view of the scheme (i.e., scheme cashflows being greater than expected) is both higher and occurs seven years after the peak of the "upside" risk (in 2043 compared with 2036). We conjecture that this is partly because the cashflows from the scheme are bounded below by zero, so scenarios with high mortality rates run the liabilities off quicker than the scenarios with low

mortality rates.

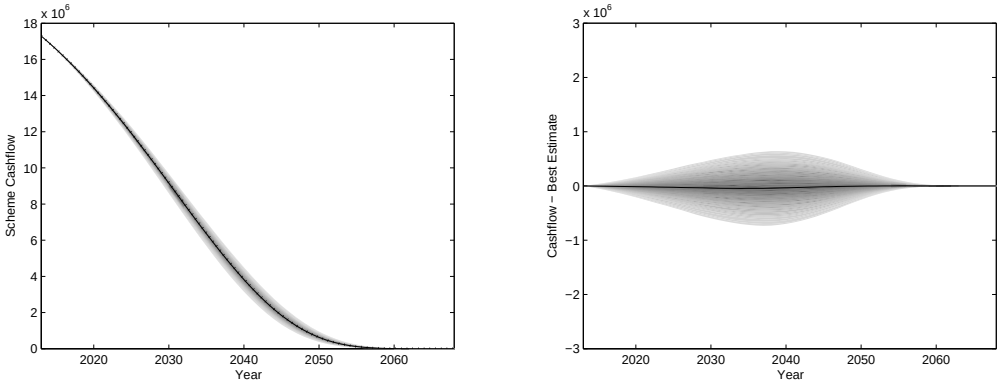
Furthermore, $\alpha_x^{(\Delta)}$ is related to the projected scheme mortality rate exponentially (i.e., $\mu_{x,t}^{(S)} \propto \exp\left(\alpha_x^{(\Delta)}\right)$). This means that, even if the level basis risk in $\alpha_x^{(\Delta)}$ is symmetrical around its best estimate, the impact on the scheme mortality rates will be asymmetrical. In addition, the asymmetry will be influenced by the interaction between the basis in the scheme, the individual scaling factors and the amount of pension in payment for individuals. This underscores the point that it is important to take account of as many mortality-related factors as possible when modelling a pension scheme, since they are likely to interact in a highly complicated fashion.

Looking at the probabilities of the scheme receiving a positive net cashflow, we find that level basis risk is a shorter term risk factor than systematic longevity risk. Although $P(t)$ is very small for $t \leq 8$, meaning that level basis risk on its own is unlikely to result in a positive net payment for the scheme, it grows rapidly after eight years (in contrast to systematic longevity risk which only gave significant probabilities of a positive net payment after around eleven years). This is not surprising, since uncertainty in the level of mortality rates will affect the scheme cashflows immediately, as opposed to needing to be compounded as in the case of systematic longevity risk. Hence, we find that more of the risk transfer provided by a bespoke longevity swap in the short term will be due to level basis risk than systematic longevity risk, in addition to level basis risk being greater overall.

However, we note that our estimates of level basis risk have been derived from a statistical analysis of the scheme itself. In practice, the involvement of a life insurer in modelling the best estimate of the cashflows will give the scheme access to more data and more sophisticated techniques for evaluating the level basis, and this should help reduce the level basis risk from the magnitudes found in this study. This means that the process of entering into a longevity swap (or, indeed, a buy-in) can help reduce the mortality and longevity risks the scheme faces in excess of just the insurance value of the contract.

10.6.4 Trend basis risk

The impact of trend basis risk on the projected cashflows of the scheme is shown in Figure 10.9, with a standard deviation for the present value of £8.0m or approximately 2.8% of the best estimate liability value of £290.5m. This suggests that trend basis risk



(A) Scheme cashflows (B) Net cashflows

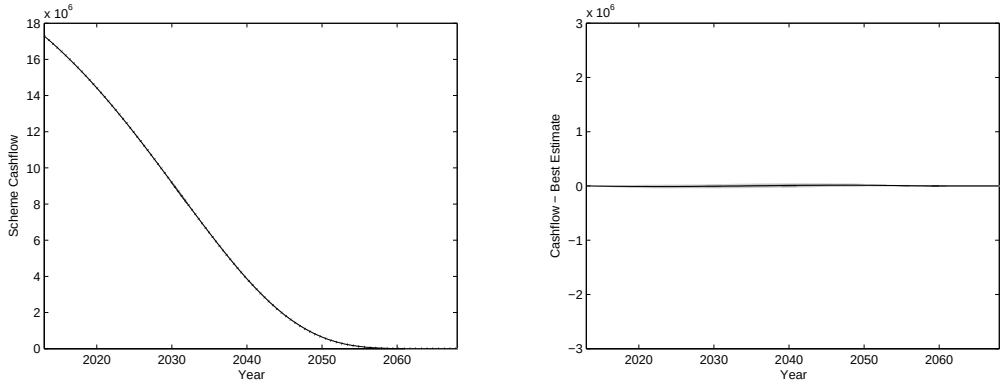
FIGURE 10.9: Impact of trend basis risk on projected scheme cashflows

is a moderately sized risk compared with the other risks modelled in this study.

As discussed in Section 10.4.2, measuring trend basis risk is very difficult, due to the fact that estimation of the trend basis is itself very difficult. Although we have used a procedure which we believe gives results that are biologically reasonable and consistent with those in other studies, there is substantial model risk in our approach. In addition, trend basis risk is a key concern for many pension scheme trustees, and anecdotally is believed to be a major limiting factor holding back the development of a market in index-based longevity swaps.

However, we believe that the overall impact of trend basis risk we find is reasonable. In particular, we note that our findings are broadly consistent with the results of Villegas and Haberman (2014), which found that allowing for trend differences in different socio-economic groups makes less than a 1% difference in the present value of annuity values at higher ages. Villegas and Haberman (2014) suggested that *"assuming the absence of improvement differentials in mortality is in principle reasonable in the valuation of annuities"*, which is consistent with not allowing for trend basis risk in the best estimate assumption.

We also observe, from looking at $P(t)$, that trend basis risk is a comparatively long-term risk for the scheme, for similar reasons as systematic longevity risk. Furthermore, because trend basis risk is of smaller magnitude than systematic longevity risk, we find that only a small component of the insurance value of the swap is in respect to trend basis risk even in the long term.



(A) Scheme cashflows (B) Net cashflows

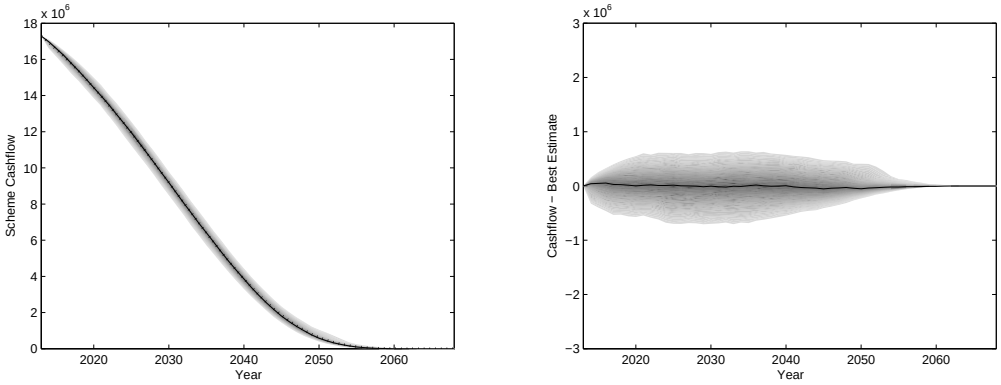
FIGURE 10.10: Impact of uncertainty in individual scalings on projected scheme cashflows

10.6.5 Uncertainty in the individual income-related scaling factors

The projected cashflows of the scheme and the swap allowing for uncertainty in the individual income-related scaling factors are shown in Figure 10.10.The standard deviation in the present value of the cashflows due to the uncertainty in the individual income-related scaling factors is only £0.6m (0.2% of the best estimate present value).

This result may be surprising, given the impact that the individual scaling factors made on the best estimate of the cashflows in Section 10.5. This explanation may, in part, be because the approach we use to allow for uncertainty in the individual scaling factors allows for more uncertainty for individuals with high scaling factors than for those with low scaling factors, due to the use of the lognormal distribution. Individuals with high scaling factors are those with the smallest amount of pension in payment, and so this uncertainty makes comparatively little impact on the projected scheme cashflows. However, we should be aware that the approach we have used to allow for uncertainty in the income-related scaling factors is quite informal and, while we believe the magnitude of the impact is reasonable, more research is required in order to fully understand the potential uncertainty in any method of assigning individual scaling factors for mortality.

Despite this proviso, our results indicate that the most important factor in the analysis is the relationship between higher income and lower mortality rates: adding uncertainty to the latter reduces the strength of the relationship but does not eliminate it. To illustrate, it is important to recognise that high-income pensioners have lower than average mortality rates, but the amount lower (e.g., whether their mortality rates are 30% or 35%



(A) Scheme cashflows (B) Net cashflows

FIGURE 10.11: Impact of idiosyncratic risk on projected scheme cashflows

lower) is less important when making projections for the scheme. In other words, it is important to build into the model the assumption of lower mortality rates for high-income pensioners, but the precise quantum of the relationship is less important.

10.6.6 Idiosyncratic risk

The projected cashflows of the scheme and the swap allowing for idiosyncratic risk in the timings of individual deaths are shown in Figure 10.11. As can be seen, idiosyncratic risk is an important risk in the context of the scheme, with a standard deviation for the present value of £6.8m or c. 2.3% of the best estimate value. This might be surprising given that there are 2,000 members of the scheme and the results of Aro (2014) and Donnelly (2014) apparently suggest that idiosyncratic risk decreases rapidly with scheme size.

However, both of these studies assumed that all members received the same amount of benefit in payment, and, therefore, weighted all lives equally. We find that the diversification of idiosyncratic risk is less effective when the lives are not equally weighted, especially in the case when the amount of pension in payment differs greatly between scheme members, as it does here. This means that diversification and the application of the law of large numbers is less effective. This is reinforced by the assumption that higher-income pensioners (with greatest weight) have a lower probability of death in any given year. Accordingly, we believe that most pension schemes are still subject to considerable idiosyncratic risk, especially in regard to the members with the largest amounts of pension in payment.

It is also interesting to note that the pattern of variability due to idiosyncratic risk is unlike that for the other risks, with a confidence interval for the cashflows which is

relatively large after only a couple of years and then stays at around this width for decades. This is because most of the idiosyncratic risk will be associated with the timing of the death of a relatively small number of individuals with large pensions, which is a risk that does not grow with time. Therefore, in the short run, this idiosyncratic risk is likely to be the dominant risk a longevity swap provides insurance against.

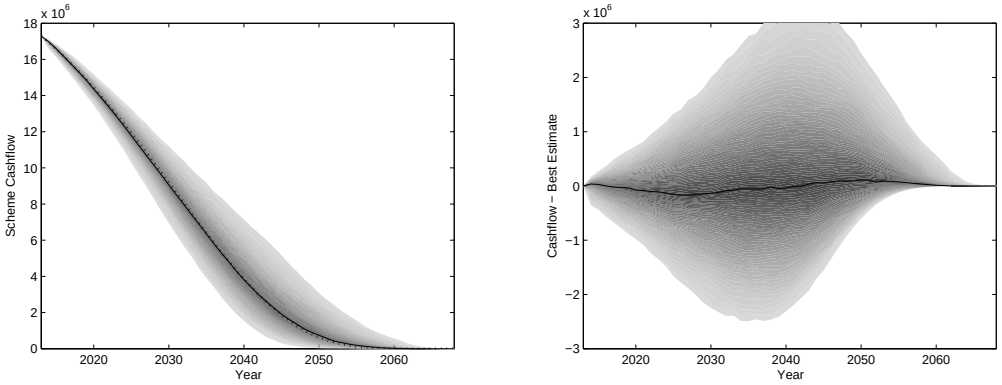
10.6.7 Summary

Figures 10.7b, 10.8b, 10.9b, 10.10b and 10.11b give some indication of the relative importance of the different mortality and longevity risks in the scheme. A summary of the information for each individual risk factor, along with the total impact all risk factors have in aggregate for the scheme, is shown in Table 10.3 and Figure 10.13. We note that these risks are largely independent of each other and, hence, that the total variance of the present value of the scheme cashflows is roughly equal to the sum of the variances for each individual risk factor. Of course, this implies that the standard deviations of the present values are not additive, as shown by the "Diversification" item in Figure 10.13.

Risk	$\mathbb{S}t\mathbb{D}ev(PV)$ (£m)	$P(2022)$	$P(2032)$
Systematic longevity risk	13.6	1%	32%
Parameter uncertainty	0.5	0%	0%
Level basis risk	27.4	9%	36%
Trend basis risk	6.8	0%	17%
Uncertain income-related scaling factors	0.6	0%	0%
Idiosyncratic risk	10.8	6%	20%
All risks	32.8	19%	38%

TABLE 10.3: Impact of different mortality and longevity risks on the present value of scheme cashflows and the probability of a positive net payment from the swap

As can be seen, the most important risk factor in terms of the standard deviation of the present value of the scheme cashflows is level basis risk, with systematic longevity risk, trend basis risk and idiosyncratic risk as the next most important risks. We also see that the swap provides significant value as an insurance policy in the shorter term, with a 19% probability of the scheme receiving a positive net cashflow in the tenth year since the inception of the swap, which doubles to 38% over a 20-year time horizon. After ten years, the main contributors to the probability of a positive net cashflow are level basis risk and idiosyncratic risk, since these impact the cashflows immediately. However, over the longer term, both systematic longevity risk and trend basis risk also provide significant contributions to the insurance value of the swap.



(A) Scheme cashflows (B) Net cashflows

FIGURE 10.12: Impact of all mortality and longevity risks on projected scheme cashflows

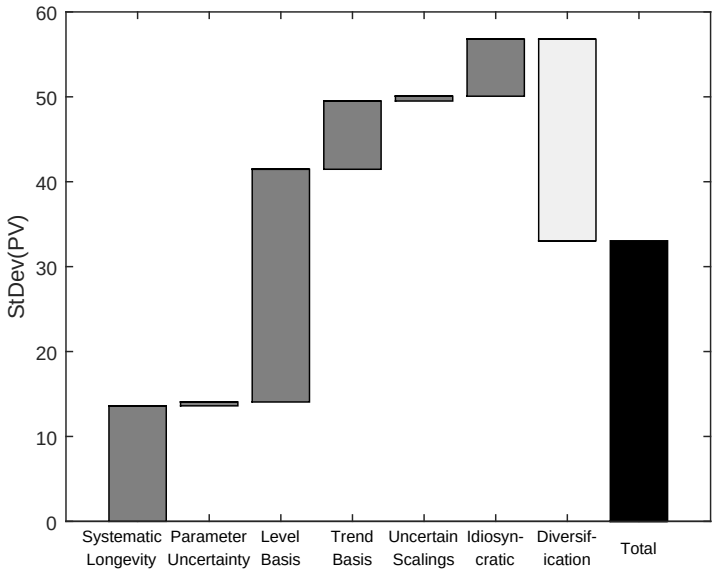


FIGURE 10.13: Contribution of each risk factor to total mortality and longevity risks for the scheme

It is also interesting that these result go a long way to explaining why pension scheme trustees prefer bespoke to index-based longevity swaps. An index-based swap only hedges the systematic component of the longevity and mortality risks present in the scheme, which is a minority of the total risk. Since these other risks are unrewarded (unlike investment risk, where pension schemes expect to earn a premium for holding the risk), it makes sense to transfer them as well as the systematic longevity risk via a bespoke longevity swap, rather than continue to hold and manage these risks internally. This is discussed further in Section 10.7.

We also note that level basis risk and idiosyncratic risk are both, in theory, reduced by

increasing the size of the scheme. This would allow the scheme to diversify the idiosyncratic risk and obtain more precise parameter estimates for the level basis. This is the most common explanation for why level basis risk and idiosyncratic risk are often overlooked in studies of risk in pension schemes. However, in practice, increasing the size of the scheme would have to be achieved by either enrolling new members into the scheme or by merging different occupational scheme together. Enrolling new members into the scheme would require the support of the corporate sponsor of the pension scheme, and is unlikely to occur, especially now that most defined benefit pension schemes in the UK are closed to new members. Furthermore, merging schemes is administratively complex and requires the consent of numerous stakeholders which may be difficult to obtain, especially as most schemes are in run-off with a view to being bought out eventually. In practice, therefore, a pension scheme's ability to increase its size is limited and, hence, level basis risk and idiosyncratic risk remain important risks for the majority of pension schemes.

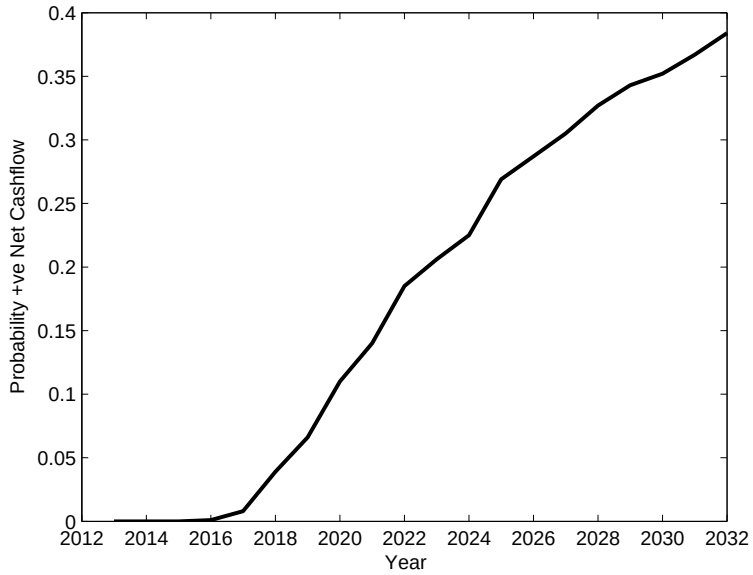


FIGURE 10.14: Probability of a positive net cashflow to the scheme

Figure 10.14 shows $P(t)$, the probability of the scheme receiving a positive net cashflow from the swap (i.e., the stochastically projected cashflows of the scheme are greater than 104% of the best estimate cashflows) in the first twenty years after inception allowing for all the different risks models. As expected, this probability grows monotonically with time. However, a scheme entering into a longevity swap would need to wait around five years before there is a significant probability of receiving a positive net cashflow from the swap. This might make it harder for the trustees of the scheme to justify entering into the swap to other stakeholders in the pension scheme, such as the corporate sponsor. This is because in the early years, the scheme is making significant net payments to

the swap provider but receiving little protection against mortality and longevity risks in return.

Nevertheless, the bespoke longevity swap still has substantial value as an insurance policy, however, since the probability of the scheme receiving a positive net cashflow from it rises rapidly to around 40% after around twenty years. However, many of the schemes entering into longevity swaps have a phased de-risking plan, of which longevity risk transfer is just one stage and the last step is a full buy-out. Depending on the timescales of the plan, it may not make sense to purchase stand-alone insurance against mortality and longevity risks that could take decades to materialise. Therefore, the timescales of the trustees' plans regarding the de-risking of the scheme will influence whether or not they consider a bespoke longevity swap to offer sufficient value as a long-term insurance policy to be justifiable.

10.7 Conclusions

The market for bespoke longevity swaps has grown rapidly in the UK and shows no signs of slowing down in the near future. In this study, we investigate the impact of various mortality and longevity risks in the context of a stylised pension scheme, in order to assess the potential for a longevity swap to transfer these risks from the scheme.

On balance, we believe that bespoke longevity swaps are valuable risk management tools for pension schemes, since the risks that they transfer are large relative to the size of the projected cashflows. However, we find that the risk factors which are often given as the main reasons for entering into a longevity swap, such as systematic longevity risk or trend basis risk, do not represent the majority of the risk being transferred. In contrast, other major risks transferred, such as level basis risk and idiosyncratic mortality risk, are not often given by pension scheme trustees as motivations for entering into a longevity swap. This may be because it is assumed that, for a large scheme, these factors can be either be measured accurately or diversified away. Since the stylised scheme investigated in this study is comparatively large by UK standards, it may appear surprising that we still find that idiosyncratic risk and level basis risk are still the largest risk factors present in the scheme. However, there are several reasons why they remain more significant than might be expected, even for a comparatively large pension scheme.

In respect of idiosyncratic risk, the heterogeneity of the membership in most pension schemes means that the idiosyncratic risk is not diversified away as effectively as might be believed. A typical pension scheme will have a few members with large pensions who contribute most to the liabilities. A great deal of risk attaches to these individuals, which is not significantly diversified by increasing the number of scheme members with relatively low amounts of pension in payment. In addition, since these individuals are likely to live longest (due to the positive relationship between income and life expectancy), this risk is magnified.

In respect of level basis risk, we note that although the level basis itself is frequently estimated for pension schemes via an experience study or by considering the occupation of scheme members, there has usually been less attention paid to quantifying the uncertainty in this estimate. Performing an experience study but not adequately measuring the uncertainty in its findings gives an illusion of certainty, yet the uncertainty in the estimate of the basis is likely to be highly significant. This is due to the relatively small number of members in most pension schemes and the often short periods of observation used for a typical experience study. Failing to quantify the uncertainty in the estimate of the basis has the effect of substantially underestimating the level basis risk, and hence the total risk, in the scheme. As discussed in Section 10.6.3, the involvement of a life insurer can help reduce level basis risk simply by using more data and more sophisticated techniques to estimate the level basis. However, the uncertainty in these estimates should still be quantified in order to accurately measure and manage the mortality and longevity risks in the scheme.

In contrast, systematic longevity risk and trend basis risk, which are often given as the major risk factors pension scheme trustees are trying to transfer in a longevity swap, are smaller than might be expected. This is mostly because these risks take substantial time to emerge and dominate the uncertainties in current mortality rates, by which time many of the scheme members will have died. These risks are, therefore, likely to be more important when considering the deferred members of a pension scheme. However, deferred members are usually not covered by a bespoke longevity swap, partly due to the additional longevity risk for these members but also because they often retain options as to what benefits they will receive after retirement (creating additional and unquantifiable uncertainty).

These results also have significant implications for the emergence of a market in index-based longevity swaps, which would only transfer the systematic longevity risk. Index-based swaps would, therefore, appear to offer comparatively poor risk transfer in the majority of realistic situations for pension schemes and so it is not surprising that none have been transacted with pension schemes to date.[28] However, index-based swaps can still provide efficient transfer of systematic longevity risk for the very largest pension schemes to life insurers, between life insurers or between life insurers and the capital markets. In these situations, an index-based transaction could be conducted at lower cost than a bespoke swap, and life insurers or very large pension schemes are better able to diversify and manage the remaining mortality and longevity risks present in the provision of annuities and pension benefits.

Furthermore, we find that the contribution of trend basis risk to the total risk for the scheme is relative modest, in contrast to level basis risk. This result should be treated with some caution, since we also find that it is very difficult to quantify trend basis risk objectively, which is also consistent with the findings of Haberman et al. (2014), i.e., that assessing trend basis risk in a pension scheme is not practical for schemes with fewer than 25,000 members or less than eight years of reliable experience data. However, we are confident that the magnitude of the risk we find is reasonable. The difficulty in measuring the importance of trend basis risk may be one reason that it is of greater concern for many trustees of pension schemes than many of the better understood risks, since it is an "unknown unknown". Nevertheless, our results should provide some comfort that trend basis risk is manageable for most schemes.

We also find that allowing for individual scaling factors to account for income-related mortality effects is very important in projecting the best estimate of the scheme cash-flows. However, while making a broad allowance for the relationship between income and mortality is important, we find that the precise quantum of the relationship is less important. In practice, this means that results obtained using publicly available data sources on the relationship between mortality rates, income and location are not too dissimilar from those obtained from more expensive postcode analyses which make use of proprietary data. We, therefore, hope that this enables smaller schemes to allow for this relationship without incurring the additional costs associated with a full postcode analysis.

[28] We note, however, that the transaction between Pall (UK) and JP Morgan in 2011 used standardised q-forwards to hedge longevity risk for deferred members, see Blake et al. (2013).

Our findings indicate that entering into a bespoke longevity swap is more advantageous for smaller pension schemes, since the largest mortality and longevity risks are those which could, in principle, be diversified. This conflicts with the fact that most of the deals to date have involved comparatively large schemes (larger than the stylised scheme used in this study). In addition, to provide a greater transfer of risk, smaller longevity swap transactions would also give greater scope for individual underwriting of scheme members, which has only occurred in buy-out and buy-in deals to date (see Blake and Harrison (2013)). The use of individual underwriting would reduce the potential uncertainty in the individual scaling factors and the overall level basis for the scheme. However, individual underwriting is less practical for very large schemes and so has not been a feature of the longevity swaps transacted to date.

If the market for bespoke longevity swaps moves to targeting smaller occupational pension schemes in future, we believe that longevity swaps (as well as buy-ins and buy-outs) will be largely beneficial in managing longer-term longevity risks in the economy. The reasons for this are twofold. First, it will help with the transfer of longevity risk from the lightly regulated occupational pension scheme sector to the better capitalised insurance sector. The recent financial crisis underscored the importance to the stability of the economy of adequate capital being provided to support risks. Hence, transferring risks from underfunded occupational pension schemes to well-capitalised insurance companies is likely to improve overall economic stability in respect of unforeseen longevity shocks. Second, buy-outs, but-ins and longevity swaps perform an important role in the aggregation of longevity risk, since an insurer transferring risk from many different schemes will be able achieve the scale needed to diversify idiosyncratic risk and obtain more certain estimates of any level basis. Furthermore, most insurers have access to better tools to measure the scheme-specific and individual-specific mortality factors than a typical pension scheme, meaning that these risks can potentially be better managed in the insurance sector.

However, this aggregation process will still leave the insurer exposed to undiversifiable systematic longevity risk. Accordingly, we see the management of longevity risk occurring in a two-stage process, with the full range of mortality and longevity risks first being transferred to the insurance sector, and then the systematic longevity risk being transferred onwards to investors in the capital markets who would otherwise not be exposed to it. In order to achieve this second step, it is important that longevity-linked securities indexed to national population data, such as index-based longevity swaps, exist to enable them to manage this undiversifiable risk, whilst the insurer retains those risks it

can diversify and manage.

In summary, we believe that, for many UK pension schemes, it is worth buying a longevity swap. However, it is important to fully assess the mortality and longevity risks in the pension scheme in order to determine which risks being transferred are the most important, as they may not be those typically given as reasons for entering into a swap.

10.A Scheme data generating process

To generate data for the stylised pension scheme, we start by specifying the total number of scheme members drawing pensions. We set this as 2,000, which represents a scheme in the largest 20% pension schemes in the UK according to The Pensions Regulator (2013b). This number was chosen as it would typically be expected to give total pensioner liabilities of c. £250m, which is at the lower end of the range of schemes which has been targeted for longevity swaps to date.

We then populate the scheme with members according to certain criteria, in order to give a realistic membership profile. First, we assume that each member being populated has an equal probability of being male or female. Next, we assign the member a retirement date randomly, with these dates distributed between zero and 25 years ago. The probability density of this distribution is assumed to decrease linearly to be zero for retirement beginning 25 years previously, i.e., members are more likely to have retired recently, consistent with a scheme which is maturing. From this, member ages can be calculated based on a retirement age of 65, which is typical for many schemes in the UK. This gives pensioner ages between 65 and 90, where 90 is the maximum age in the SAPS data used in our analysis and so makes a suitable choice for the cut off in retirement dates.

This procedure populates a large number of members who would have retired from the scheme. Clearly, not all members who retired would have survived to the present day. To allow for this, we need to allow approximately for mortality between retirement and the current date. To do this, we use mortality rates for the UK national population in 2011 to estimate the survival probability of the member from retirement to the present.[29] The survivorship of an individual is then modelled as a Bernoulli random variable with

[29] This implicitly assumes that mortality rates have remained constant over this period for simplicity. This assumption overstates the survivorship of individuals as it will not reflect the mortality improvements over the retirement period experienced by real pensioners, and so result in our stylised scheme being slightly older than is typical.

the estimated survival probability. Only members who are currently alive are included in the data, and we cease generating new members when we have 2,000 living individuals.

For each living member, we then allocate them to an income percentile at retirement. This is determined by a uniform random variable between zero and one, which we then use to assign an individual scaling factor in the manner described in Section 10.4.2. From this income percentile, we generate a pension amount for the member based on an assumed distribution for income at retirement today. Pension amounts are assumed to be distributed according to the Pareto distribution, which has often been used to model the distribution of income within a population. This distribution is capable of replicating the observed levels of inequality in income between individuals, due to the long tail of the observed income distribution where a small number of members (typically former directors of the sponsoring employer) have extremely large pensions relative to the median.

We assume that the Pareto distribution used to generate the pension amount has a threshold of £3,000 and a scale factor of 1.43. This threshold was chosen to reflect the typical amount of pension below which members can trivially commute their entire pension benefits to cash at retirement, and therefore leave no residual liability with the scheme. The scale factor has been chosen based on the "10/50" rule discussed in footnote 9, i.e., that 10% of the scheme members receive 50% of the pension in payment. The Pareto distribution can be calibrated to produce distributions of pensions consistent with any rule in the form "X% of the scheme membership receive Y% of the pension in payment" by choosing a scale factor such that

$$\alpha = \frac{\ln X}{\ln X - \ln Y}$$

This income distribution is expressed as an amount of pension if they retired in 2011. However, pensioners who retired before this date are likely to have lower pensions in payment today because they would have had lower salaries (in real terms) when they retired and would have received pension increases linked to price inflation rather than salary inflation (which is usually higher). Therefore, the pension amount if the member retired today is converted to a pension amount currently in payment by adjusting for real salary increases, which are assumed to be equal to 0.5% p.a., i.e.,

$$\text{Pension in payment} = \text{Pension if retired today} \times 1.005^{\text{Age}-65}$$

Together, these four variables - sex, age, scaling factor and pension in payment - are allocated to each surviving member of the scheme for use in the projections in this study, and are summarised in Figures 10.2 and 10.3 in Section 10.3.

Part IV

Forward Mortality Models

Chapter 11

Forward Mortality Rates in Discrete Time I: Calibration and Securities Pricing

11.1 Introduction

Many users of mortality models are interested in using them to place values on longevity-linked liabilities and securities. Modern regulatory regimes require that the values of liabilities and reserves are consistent with market prices (if available), whilst the gradual emergence of a traded market in longevity risk needs methods for pricing new types of longevity-linked securities quickly and efficiently. These needs have spurred the development of increasingly sophisticated models of mortality rates.

Cairns et al. (2006b) pointed out that the majority of mortality models that have been proposed are models of the mortality hazard rate, which is analogous to the short rate of interest. By analogy with interest rate models, Cairns et al. (2006b) developed formally the concept of "mortality forward rates", which was extended in Miltersen and Persson (2005). However, the idea of forward mortality rates has a long history, indeed Milevsky and Promislow (2001) pointed out that *"the traditional rates used by actuaries are really 'forward rates' exactly analogous to a forward interest rate implied by existing bond prices"*.

Such forward mortality rates could be used to price longevity-linked securities, in the same fashion as forward interest rates are used to value cashflows dependent on future

interest rates. Therefore, a number of models for forward mortality rates have been proposed to date which build upon the theory of forward interest rates. These have included the models of Barbarin (2008), Bauer et al. (2008) and Tappe and Weber (2013), which adopted the Heath-Jarrow-Morton framework used for interest rates in continuous time, and the model of Zhu and Bauer (2011a,b, 2014) which adopted a semi-parametric factor approach in discrete time. An alternative approach, developed in Olivier and Jeffrey (2004), Smith (2005) and Cairns (2007), also works in discrete time but uses gamma-distributed random variables to update a forward mortality surface that is initially assumed.

However, it is important not to over-extend the analogy between interest rates and mortality rates, as the two are fundamentally different processes. Most obviously, the forward interest rate curve at any instant depends only upon term, whilst forward mortality rates will exist across a surface of ages and years. Mortality rates typically also increase exponentially with age, unlike interest rates which are typically bounded as term increases. More fundamentally, the analogy between survivorship under a force of mortality and discounting under a force of interest, whilst mathematically appealing, is not exact, since mortality will affect the actual amount of any cashflow payable (say, in an annuity or life assurance contract) in a way that discounting does not. We therefore do not believe that simply taking existing models which work well for forward interest rates and applying them directly to mortality rates is appropriate.

In addition, we must be able to calibrate a model of forward mortality rates to the small number of longevity-linked securities in existence. This means that models which start by assuming the existence of sufficient market prices to define a forward mortality surface (such as those based on the Heath-Jarrow-Morton framework) and then define the dynamics of this surface are not practical. This approach is inherited from the interest rate markets, where liquid markets in bonds across the whole of the relevant term structure can provide such information. Unfortunately, this simply does not hold for the market in longevity-linked securities, and will not hold for the foreseeable future.

Instead, we propose a new approach, which is described in two studies, of which this is the first. Our approach starts from the historical data on the observed mortality rates, i.e., the observed force of mortality which is analogous to the short rate of interest. Building on the dynamics of models of the observed force of mortality, we can recast them in the form of models of forward mortality rates and then use a change of measure to incorporate whatever market information is available. This approach ensures

that the dynamics of the forward mortality surface are consistent with those observed for the force of mortality, including features such as "cohort effects" which are unique to mortality rate models, and which helps to ensure demographic significance.[1]

We begin our analysis in this paper in Section 11.2.1 with models of force of mortality from the age/period/cohort (APC) family, which have been specifically constructed in order to capture the dynamics of mortality parsimoniously and with demographic significance. APC mortality models are considered in detail in Chapters 2, 3 and 4 and encompass a broad class of existing and popular models of the force of mortality, such as the Lee-Carter (Lee and Carter (1992)), Cairns-Blake-Dowd (Cairns et al. (2006a)) and classic APC (Hobcraft et al. (1982)) models, as well as many of the extensions of these models (see Chapter 5 for examples). We then develop the mathematical framework required to convert any APC model of the force of mortality into a model of the forward mortality surface in Section 11.2.2 and Section 11.2.3. In Section 11.2.4, we use the dynamics of the period and cohort parameters observed in the historical data to define a forward surface of mortality rate. This enables consistent modelling of both the short and forward mortality rates, and so avoids any inconsistencies between the two.

Section 11.3 then builds on this by transforming the forward mortality rate surface, using the Esscher transform, from a measure consistent with the "real-world" process observed in the historical data to one consistent with market prices. These "market-consistent" forward mortality rates are then used to price various longevity-linked securities. Finally, Section 11.4 concludes.

The approach established in this chapter is extended in our second study, Chapter 12, which analyses how the forward surface of mortality can be updated dynamically. This enables the forward mortality rate framework developed in this study to be used for managing longevity risk in a life assurance book or in a portfolio of longevity-linked securities.

[1] Demographic significance is defined in Chapter 2 as the interpretation of the components of a model in terms of the underlying biological, medical or socio-economic causes of changes in mortality rates which generate them.

11.2 Forward mortality rates in discrete time

11.2.1 Age/period/cohort models of the force of mortality

In Chapter 2, we discussed discrete-time mortality models of the form

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{t-x} \tag{11.1}$$

where

- we have historical data for ages, x, in the range $[1, X]$ and periods, t, in the range $[1, \tau]$ and therefore observations of cohorts born in years, y, in the range $[1 - X, \tau - 1]$;
- $\eta_{x,t} = \ln(\mu_{x,t})$ is the log-link function which connects the Poisson distributed death counts, $D_{x,t}$, to the proposed predictor structure;
- α_x is a static function of age;
- $\kappa_t^{(i)}$ are period functions governing the evolution of mortality with time;
- $\beta_x^{(i)}$ are age functions modulating the impact of the period function dynamics over the age range;[2] and
- γ_y is a cohort function describing mortality effects which depend upon a cohort's year of birth and follow that cohort through life as it ages.

Defining $\boldsymbol{\beta}_x = \left(\beta_x^{(i)}, \quad \ldots \quad \beta_x^{(N)}\right)^\top$ and $\boldsymbol{\kappa}_t = \left(\kappa_t^{(i)}, \quad \ldots \quad \kappa_t^{(N)}\right)^\top$, we can re-write Equation 11.1 as

$$\eta_{x,t} = \alpha_x + \boldsymbol{\beta}_x^\top \boldsymbol{\kappa}_t + \gamma_{t-x} \tag{11.2}$$

In this study, we will use the log-link function $\eta_{x,t} = \ln(\mu_{x,t})$. In Chapter 2, we discussed how this is appropriate if the death count at age x and time t is a (conditionally independent) Poisson random variable, $D_{x,t} \sim Po(\mu_{x,t} E^c_{x,t})$, where $E^c_{x,t}$ are central exposures to risk. This is preferred over the alternative choice of the logit-link function and binomially distributed death counts due to the distributional properties of the forward

[2] These can be non-parametric in the sense of being one fitted without imposing any a priori shape for the function across ages, or be parametric in the sense of having a specific functional form, $\beta_x^{(i)} = f^{(i)}(x; \theta^{(i)})$ selected a priori. Potentially, parametric age functions can have free parameters $\theta^{(i)}$ which are set with reference to the data.

mortality rates, as discussed in Section 11.2.3.

This structure defines the class of age/period/cohort (APC) mortality models and is very flexible. Many of the most common mortality models fit into this structure, for instance, the benchmark Lee-Carter (LC) model of Lee and Carter (1992), the cohort extension to this denoted H1 in Haberman and Renshaw (2009), the Cairns-Blake-Dowd (CBD) model of Cairns et al. (2006a) and many of its extensions in Cairns et al. (2009), the Plat model of Plat (2009a) and the model of Börger et al. (2013). In Chapter 5, we describe a "general procedure" for constructing bespoke models within this class which are tailored to the structure within a given dataset.[3] It is, therefore, appropriate to use this class of models of the force of mortality as the starting point for defining the forward mortality surface, as discussed below.

11.2.2 Defining forward mortality rates

In a discrete-time framework, the force of mortality, $\mu_{x,t}$, at age x and time t is assumed to be constant over each age and year, i.e.,

$$\begin{aligned} \mu_{x+\xi,t+\tau} &= \mu_{x,t} \\ x,t &\in \mathbb{N} \\ \xi,\tau &\in [0,1) \end{aligned} \tag{11.3}$$

Therefore, the one-year survival probability from age x at time t to age $x+1$ at time $t+1$, $p_{x,t}$,[4] is equal to $p_{x,t} = \exp(-\mu_{x,t})$. If we further assume that survival in each year is conditionally independent, this implies

$${}_tp_{x,\tau} = \prod_{u=1}^{t} p_{x+u,\tau+u} = \exp\left(-\sum_{u=1}^{t} \mu_{x+u,\tau+u}\right) \tag{11.4}$$

where ${}_tp_{x,\tau}$ is the survival probability of an individual from age x at time τ to age $x+t$ at time $\tau+t$.[5] If $\tau+t$ lies in the future, ${}_tp_{x,\tau}$ will be a random variable, as future values of the force of mortality will be subject to systematic mortality risk.

[3] The forward mortality framework described in this study is not significantly affected if the cohort parameters are modulated by an age function, $\beta_x^{(0)}$, as in the model of Renshaw and Haberman (2006). However, for simplicity and the reasons discussed in Chapter 2, we do not consider such models in this study.

[4] $p_{x,t} = 1 - q_{x,t}$, the one-year probability of death.

[5] ${}_0p_{x,\tau} = 1$ trivially.

To define the structure of forward mortality rates, we assume that the fundamental longevity-linked security[6] of interest, from which all other longevity-linked securities can be constructed, is the "longevity zero".[7] A longevity zero is defined in Blake et al. (2006) as a zero-coupon bond which pays out a principal at a future time, dependent on the survivorship of a suitably large cohort (to reduce the idiosyncratic risk in the estimation of survival rates) over the term of the bond.[8] Therefore, a t-year longevity zero at time τ would have price

$$\text{Price}(t,\tau) = B(\tau,\tau+t)\mathbb{E}^{\mathbb{Q}}{}_{\tau}\,{}_{t}p_{x,\tau}$$

where $B(\tau,\tau+t)$ is the time τ price of a t-year zero coupon bond paying one unit at maturity, and where the expectation is defined under some "market-consistent" measure, $\mathbb{Q}$ (to be discussed in Section 11.3).[9]

In doing so, we have implicitly assumed that the longevity risk is independent of the other financial risks in the market, such as interest rates and inflation, in both the real-world measure, $\mathbb{P}$, and the market-consistent measure, $\mathbb{Q}$. This is in common with the majority of studies, such as Cairns et al. (2006b) and Bauer et al. (2008) and with the available evidence to date, as discussed in Loeys et al. (2007). Although there may be some situations where longevity risk is not independent of other financial risks in the real-world measure, as in the examples of Miltersen and Persson (2005), we believe that these situations are relatively extreme and are better considered by scenario analysis rather than through a stochastic model. Furthermore, Dhaene et al. (2013) show that independence between longevity risk and financial risks in the real-world measure does not automatically ensure independence in the market-consistent measure. However, more complicated models are required in order to allow for any dependence between longevity and investment risks, which require more market information for calibration. Therefore, we believe that the assumption of independence between longevity risk and other financial risks is necessary and justifiable at this early stage of development of the longevity risk market.

[6] In this paper, we use the term "security" to refer to any tradable financial contract, and so also include derivative securities such as forwards and options in this definition.

[7] Longevity zeros were also used to define forward mortality rates in Barbarin (2008) for use in a Heath-Jarrow-Morton framework and in Cairns (2007) and Alai et al. (2013) to develop extensions of the Olivier-Smith model.

[8] It is important that the security used to define the forward mortality rates depends purely on the systematic longevity risk, rather than the idiosyncratic time of death of any individual lives, in order to avoid the potential for conflicting definitions of the forward rates described in Norberg (2010).

[9] We adopt the convention that the subscript on operators $\mathbb{E}_{\tau}(.)$, $\mathbb{V}ar_{\tau}(.)$ or $\mathbb{C}ov_{\tau}(.)$ denotes conditioning on the information available at time τ, i.e., $\mathcal{F}_{\tau}$.

We define

$$
\begin{aligned}
{}_tP^{\mathbb{Q}}_{x,\tau}(\tau) &= \mathbb{E}^{\mathbb{Q}}{}_\tau \, {}_tp_{x,\tau} \\
&= \mathbb{E}^{\mathbb{Q}}{}_\tau \exp\left(-\sum_{u=1}^{t} \mu_{x+u,\tau+u}\right)
\end{aligned}
\tag{11.5}
$$

In this, ${}_tP^{\mathbb{Q}}_{x,\tau}(\tau)$ are the market-consistent forward survival probabilities, i.e., the *"market's best view"* (in the words of Miltersen and Persson (2005)) at τ of the probability of an individual aged x at τ surviving a further t years. Mathematically, we can see that these factors are analogous to discount factors based on the prices of zero-coupon bonds. It is this analogy which has motivated much of the development of forward mortality rate models to date, which have been mainly adapted from widely used interest rate models. In continuous-time forward rate models, such as in Bauer et al. (2008), forward mortality rates are defined from Equation 11.5 as

$$
\nu^{\mathbb{Q}}_{x,t}(\tau) \equiv -\frac{\partial}{\partial t} \ln\left({}_tP^{\mathbb{Q}}_{x-t,\tau}(\tau)\right)
$$

via the analogy with forward interest rates. In a discrete time model, we modify this to define forward mortality rates as

$$
\nu^{\mathbb{Q}}_{x,t}(\tau) \equiv -\ln\left(\frac{{}_{t-\tau+1}P^{\mathbb{Q}}_{x-t+\tau,\tau}(\tau)}{{}_{t-\tau}P^{\mathbb{Q}}_{x-t+\tau,\tau}(\tau)}\right) \tag{11.6}
$$

Existing forward mortality models, such as those in Cairns (2007) and Zhu and Bauer (2011b, 2014) use similar definitions, but these studies are interested in the dynamics of the forward surface of mortality and so are interested in the behaviour of $\nu_{x,t}(\tau+1)/\nu_{x,t}(\tau)$, rather than the forward mortality rates at τ themselves (which are assumed a priori in these studies). We discuss these dynamics in Chapter 12. In contrast, this chapter is interested in the connection between the force of mortality and forward mortality rates, and so we use the definition above to give

$$
{}_tP^{\mathbb{Q}}_{x,\tau}(\tau) = \exp\left(-\sum_{u=1}^{t} \nu^{\mathbb{Q}}_{x+u,\tau+u}(\tau)\right) \tag{11.7}
$$

Comparing Equations 11.4 and 11.7, we see

$$
\exp\left(-\sum_{u=1}^{t} \nu^{\mathbb{Q}}_{x+u,\tau+u}\right) = \mathbb{E}^{\mathbb{Q}}{}_\tau \exp\left(-\sum_{u=1}^{t} \mu_{x+u,\tau+u}\right) \tag{11.8}
$$

which shows the connection between the market-consistent forward rates and the expectations of the force of mortality in the market-consistent measure.

By Jensen's inequality

$$\mathbb{E}^{\mathbb{Q}}{}_{\tau} \exp\left(-\sum_{u=1}^{t} \mu_{x+u,\tau+u}\right) \geq \exp\left(-\sum_{u=1}^{t} \mathbb{E}^{\mathbb{Q}}{}_{\tau}\ \mu_{x+u,\tau+u}\right) \tag{11.9}$$

In practice, the variation in $\mu_{x,t}$ is sufficiently small that Equation 11.9 holds approximately as an equality over almost all ages and years.[10] We therefore make the assumption that

$$\exp\left(-\sum_{u=1}^{t} \nu^{\mathbb{Q}}_{x+u,\tau+u}(\tau)\right) = \exp\left(-\sum_{u=1}^{t} \mathbb{E}^{\mathbb{Q}}{}_{\tau}\ \mu_{x+u,t+u}\right) \tag{11.10}$$

and define the forward mortality rates as

$$\nu^{\mathbb{Q}}_{x,t}(\tau) = \mathbb{E}^{\mathbb{Q}}{}_{\tau}\ \mu_{x,t} \tag{11.11}$$

Thus, the forward mortality rate at age x and year t is assumed to be equal to the expectation under the market-consistent measure of the force of mortality at the same age and year, conditional on information observed at time τ. Thus, if we can specify the dynamics of the force of mortality (in the market-consistent measure), we are able to find the forward mortality rates directly.

We define the "forward mortality surface" as the collection of forward mortality rates, $\nu^{\mathbb{Q}}_{x,t}(\tau)$ over all ages, x, and future years, t, at a given point in time, τ. In most cases, it is more natural to consider the forward mortality surface as a single object, since the individual forward mortality rates are expected to vary smoothly across ages and across future years. However, it is important to realise that the forward mortality surface is three-dimensional, defined by x, t and τ. In this study we shall consider its structure across the dimensions of x and t and how this can be determined at the observation time, τ, which is assumed to be constant. This contrasts with Chapter 12, where we discuss how the surface varies dynamically with τ.

In defining the forward mortality surface, we assume that all longevity-linked securities can be constructed from a portfolio of longevity zeros. We shall see in Section 11.3.3 that

[10] This approximation is tested numerically in Appendix 11.B.

this is trivially true in the case of longevity swaps.[11] We extend this by assuming that the value of any other longevity-linked security at time τ can be replicated as a portfolio of longevity zeros and, therefore, written as a function of the $\nu^{\mathbb{Q}}_{x,t}(\tau)$. Hence, the forward surface of mortality can be used to give consistent prices for all longevity-linked liabilities and securities.

Unfortunately, however, it is currently impossible to reliably specify the dynamics of short or forward mortality rates in the market-consistent measure, since an actively-traded market in longevity-linked securities does not currently exist. Indeed, the absence of genuine market information on the prices for any longevity-linked securities is a critical problem for all studies that seek to value the few longevity-linked securities which do exist. There have been a number of different methods proposed to overcome this and calibrate the market-consistent measure. For instance, Bauer et al. (2008) proposed using generational life tables (i.e., those which allow mortality rates to depend upon an individual's year of birth) in order to provide a forward mortality surface. However, these are updated infrequently and are not based on market information (and when used to price financial contracts, typically have margins for risk aversion added to them). Alternatively, Miltersen and Persson (2005) and Bayraktar and Young (2007) have suggested using the market for endowment assurances for calibration purposes, since these have a similar price structure to longevity zeros. Unfortunately, Norberg (2010) showed how using securities dependent on the idiosyncratic risk of individual lives, such as endowment assurances, can lead to inconsistent definitions of the forward mortality rates and so this approach is not feasible.

Instead, we propose to use the historical data to model the dynamics of the force of mortality in the "historical" or "real-world" measure, $\mathbb{P}$, using relatively simple APC mortality models, as described in Section 11.2.1. These real-world dynamics of the force of mortality can then be used to generate the forward surface of mortality in the real-world measure by using Equation 11.11. Then, in Section 11.3.1, we show how to change from the real-world to a market-consistent measure, $\mathbb{Q}$, using the Esscher transform which is calibrated using whatever (limited) market information for longevity risk is available. Thus, real-world data on historical mortality rates is used to supplement the limited market data we have, and increasing volumes of market information can be incorporated into the forward mortality surface as the market for longevity-linked securities develops.

[11] It is also true for the valuation of annuities for reserving purposes, since idiosyncratic risk is not allowed for in this context.

11.2.3 Forward APC mortality models

Combining Equations 11.2 and 11.11, we define forward mortality rates in the real-world measure, $\mathbb{P}$, as

$$\nu^{\mathbb{P}}_{x,t}(\tau) = \mathbb{E}^{\mathbb{P}}{}_{\tau} \exp\left(\alpha_x + \boldsymbol{\beta}_x^{\top}\boldsymbol{\kappa}_t + \gamma_{t-x}\right) \tag{11.12}$$

We assume that the age functions are known with certainty at time τ and therefore the uncertainty in future mortality rates comes from the projection of $\boldsymbol{\kappa}_t$ andγ_{t-x}, i.e., the forward mortality surface only allows for process risk from the projection of the period and cohort functions, in the terminology of Cairns (2000), but not parameter uncertainty or model risk. In the real-world measure, we first obtain fitted values of $\boldsymbol{\kappa}_t$ and γ_y by fitting the APC model to the historical data. We then estimate the dynamics of the time series processes for $\boldsymbol{\kappa}_t$ and γ_y from these fitted values.

If we further assume that our projected $\boldsymbol{\kappa}_t$ and γ_y are normally distributed, then $\eta_{x,t}$ is also normally distributed and consequently $\mu_{x,t}$ follows a log-normal distribution.[12] Therefore

$$\nu^{\mathbb{P}}_{x,t}(\tau) = \exp\left(\alpha_x + \boldsymbol{\beta}_x^{\top}\mathbb{E}^{\mathbb{P}}{}_{\tau}\boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x^{\top}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{E}^{\mathbb{P}}{}_{\tau}\gamma_{t-x} + \frac{1}{2}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\gamma_{t-x})\right) \tag{11.13}$$

The assumption that projected period and cohort parameters are normally distributed is in line with the majority of studies, which use standard ARIMA methods to project these parameters. If the projected period and cohort parameters are not normally distributed, however, it is unlikely that the resulting forward mortality framework would be analytically tractable. This is because the distribution of $\mu_{x,t}$ would not have the finite moments required. A number of studies have used alternative methods and distributions to make projections. These include models which allow for regime changes (Milidonis et al. (2011) and Lemoine (2014)) or trend changes (Sweeting (2011) and Chapter 6) in the processes used to project the parameters. Another approach has been to use other distributions for the innovations in the time series processes for the period or cohort functions (such as the t-distribution, the variance-gamma and the normal-inverse-gamma, which were used to model the innovations for κ_t in the Lee-Carter model in Wang et al. (2011)). In some of these cases, it may be possible to extend the forward mortality rate framework

[12] Note that, if we were using $\eta_{x,t} = \text{logit}(q_{x,t})$ in conjunction with a binomial model for the death count, then $q_{x,t}$ would follow a "logit-normal" distribution (see Frederic and Lad (2008)). Unfortunately, this is not analytically tractable and does not possess closed form expressions for the expectation. Therefore, we are unable to define a forward mortality framework in the logit-link function / binomial death count model as we can in the log-link function / Poisson death count model.

to allow for the non-Gaussian distributions. However, we do not consider alternative distributions for the projected period or cohort functions further within this study.

11.2.4 Projecting the APC model

11.2.4.1 Period functions

Since Lee and Carter (1992), the most common method used to project the period functions in an APC mortality model has been the random walk with drift. This was also used for the CBD model in Cairns et al. (2006a), the period functions in various mortality models in Cairns et al. (2011a) and Haberman and Renshaw (2011), and the first (dominant) period function in Plat (2009a).

The random walk model is attractive as it allows the period functions to be non-stationary with a variability that increases with time, giving biologically reasonable[13] projections of the force of mortality.

In Chapters 3 and 4, we discuss how projected mortality rates should not depend upon the identifiability constraints used when fitting the model to data, and therefore that we should use "well-identified" projection methods which achieve this. In the context of the random walk with drift model, this means we should project the period functions using

$$\boldsymbol{\kappa}_t = \mu X_t + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\epsilon}_t \tag{11.14}$$

where X_t is a set of deterministic functions ("trends") chosen to ensure identifiability and μ are the corresponding "drifts".[14] For example, the classic random walk with drift process has a constant trend, $X_t = 1$, with the "drift", μ, found be regressing $\Delta\boldsymbol{\kappa}_t$ on this trend. Similarly, the random walk with linear drift introduced in Chapters 4 and 6 has constant and linear trends, $X_t = \begin{pmatrix} 1, & t \end{pmatrix}^\top$, with the drifts found by regressing $\Delta\boldsymbol{\kappa}_t$ against X_t in a similar fashion.

[13] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

[14] Note, we assume that the drifts μ are known at time τ and will not be re-estimated on the basis of new information arising in the future. Therefore, the forward mortality framework described in this chapter and in Chapter 12 does not allow for "recalibration" risk as defined in Cairns (2013), i.e., the risk caused by the uncertainty in the drift. This risk is potentially substantial, as discussed in Li et al. (2004) and Li (2014). However, we leave the inclusion of recalibration risk to future work.

The random drift model in Equation 11.14 is solved to give

$$\boldsymbol{\kappa}_t = \boldsymbol{\kappa}_\tau + \mu\chi_{\tau,t} + \sum_{s=\tau+1}^{t} \boldsymbol{\epsilon}_s \tag{11.15}$$

where $\chi_{\tau,t} = \sum_{s=\tau+1}^{t} X_s$. Note that, in the simplest case where we use a classic random walk with drift to project the period functions, $X_t = 1$ and hence $\chi_{\tau,t} = t - \tau$. We assume

$$\mathbb{E}_\tau \boldsymbol{\epsilon}_t = \mathbf{0}$$
$$\mathbb{C}ov_\tau\left(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_s\right) = \Sigma \mathtt{I}_{t-s}$$

where $\mathtt{I}_{t-s}$ is an indicator variable taking a value of unity if $t = s$ and zero otherwise. This means that the innovations have zero mean and are independent across different periods, i.e., they are white noise. In addition, we assume that the innovations are normally distributed for the reasons discussed above. From Equation 11.15, we find

$$\mathbb{E}^{\mathbb{P}}{}_\tau\, \boldsymbol{\kappa}_t = \boldsymbol{\kappa}_\tau + \mu\chi_{\tau,t} \tag{11.16}$$
$$\mathbb{V}ar^{\mathbb{P}}_\tau\left(\boldsymbol{\kappa}_t\right) = (t-\tau)\Sigma \tag{11.17}$$

In an age/period mortality model without a cohort term, such as the Lee-Carter or CBD model, allowing for the uncertainty in the period functions is sufficient in conjunction with Equation 11.13, to define forward mortality rates in the real-world measure. However, more sophisticated mortality models often include cohort terms, whose analysis is considerably more complicated, as we now see.

11.2.4.2 Cohort function

Most common techniques for projecting the cohort function use standard ARIMA processes, which assume that there is a clear distinction between those cohort parameters which are estimated from historical data, which are assumed to be known, and those cohort parameters which are projected using some time series process. In the forward mortality rate framework, we can see that this would lead to a sharp discontinuity in the forward mortality surface. For many purposes, such as the valuation of longevity-linked securities and liabilities, such a discontinuity is clearly undesirable.

To illustrate this problem, consider the case where a (well-identified) AR(1) process is used to project the cohort parameters

$$\gamma_y - \beta\tilde{X}_y = \rho(\gamma_{y-1} - \beta\tilde{X}_{y-1}) + \varepsilon_y$$

where $\tilde{X}_y$ are deterministic functions corresponding to the unidentifiable trends in the cohort parameters,[15] and β are the corresponding regression coefficients (see Chapter 4). Such a process would be solved to give

$$\gamma_y = \rho^{y-Y}(\gamma_Y - \beta\tilde{X}_Y) + \beta\tilde{X}_y + \sum_{s=Y+1}^{y} \rho^{y-s}\varepsilon_s$$

for $y \geq Y$, the year of birth of the last fitted cohort parameter.[16] The variance of this process is

$$\mathbb{V}ar_\tau^{\mathbb{P}}(\gamma_y) = \begin{cases} 0 & \text{if } y \leq Y \\ \frac{1-\rho^{2(y-Y)}}{1-\rho^2}\sigma^2 & \text{if } y > Y \end{cases}$$

From Equation 11.13, we see that this would give a discontinuity in the forward mortality surface at the interface between the fitted and projected cohort parameters. Such a discontinuity would give rise to pricing anomalies and therefore cannot be permitted in a well-designed forward mortality framework. Consequently, we must use alternative processes to project the cohort parameters for use with forward mortality models.

In Chapter 6, we developed a Bayesian approach to overcome this issue. This assumes that all cohort parameters, γ_y, are random variables that are not fully observed until cohort y is fully extinct at time $y+X$. For observation times $\tau < y+X$, we have partial information based on observations of the cohort to date. This information is summarised in the estimated cohort parameters, $\overline{\gamma}_y(\tau)$, found by fitting the APC mortality model to data to time τ. From the analysis in Chapter 6, we have

$$\gamma_y|\mathcal{F}_\tau \sim N(M(y,\tau), V(y,\tau)) \tag{11.18}$$

[15] In general, these have a similar form to the deterministic functions for the period parameters, X_t, in Section 11.2.4.1.

[16] Typically, cohort parameters for the last few years of birth are not estimated due to the lack of data, for instance, see Renshaw and Haberman (2006).

where

$$\text{Þ}_{\tau-y,s} \equiv \prod_{r=0}^{s-1}(1 - D_{\tau-y+r}) \tag{11.19}$$

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}{}_{\tau}\gamma_y &\equiv M(y,\tau) \\ &= \sum_{s=0}^{\infty} \text{Þ}_{\tau-y,s}\rho^s \left[D_{\tau-y}\overline{\gamma}_y(\tau) + (1 - D_{\tau-y+s})\beta(\tilde{X}_{y-s} - \rho\tilde{X}_{y-s-1})\right]\end{aligned} \tag{11.20}$$

$$\begin{aligned}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\gamma_y) &\equiv V(y,\tau) \\ &= \sum_{s=0}^{\infty} \text{Þ}^2_{\tau-y,s}(1 - D_{\tau-y+s})\rho^{2s}\sigma^2\end{aligned} \tag{11.21}$$

for $y \leq Y$, where

$$M(y,\tau) = \rho^{y-Y}\left(M(Y,\tau) - \beta\tilde{X}_Y\right) + \beta\tilde{X}_y \tag{11.22}$$

$$V(y,\tau) = \frac{1-\rho^{2(y-Y)}}{1-\rho^2}\sigma^2 + \rho^{2(y-Y)}V(Y,\tau) \tag{11.23}$$

for $y > Y$. In this,

- D_x is the proportion of a cohort assumed to still be alive by age x;
- ρ and σ^2 are the autocorrelation and variance of the AR(1) process assumed to be driving the evolution of the cohort parameters;
- $\tilde{X}_y$ and β are the trends and drifts for the cohort parameters as defined above;[17]
- $\overline{\gamma}_y(\tau)$ are the estimates of the cohort parameters, fitted by the mortality model at time τ; and
- $\mathcal{F}_\tau$ is the total information available at time τ, including observations of the cohort parameters up to year of birth y, i.e., $\{\overline{\gamma}_\upsilon(\tau)\ \upsilon \leq y\}$.

In Chapter 6, it was shown that this framework allows the historical and projected cohort parameters to be treated consistently, without any sharp discontinuities in the uncertainty between them. It was also shown that these projections are well-identified, in the sense that they do not depend upon the arbitrary identifiability constraints made when fitting the model. In addition, it is shown in Chapter 12 that the Bayesian framework allows us to update estimates of the cohort parameters over a one-year period to proxy for the impact that new data would have on our parameter estimates, which is essential for risk management purposes. The Bayesian framework is therefore well adapted

[17] Note that the drifts, β, depend upon the arbitrary identifiability constraints chosen. In practice, we therefore impose a set of identifiability constraints such that $\beta = 0$ to simplify matters considerably.

for use in a forward mortality context, and we will use it for all APC mortality models which include cohort parameters.

11.2.5 Estimation and projection

The framework described in Sections 11.2.3 and 11.2.4 is very general and can be used in conjunction with any APC mortality model for the force of mortality. To see this in practice, we consider estimating the forward mortality rates on male data for the UK for the period 1950 to 2011 and ages 50 to 100 from the Human Mortality Database (2014) for five different APC models:

1. the Lee-Carter ("LC") model of Lee and Carter (1992);
2. the "CBDX" model discussed in Chapter 3, which extends the Cairns-Blake-Dowd model of Cairns et al. (2006a) with a static age function and uses a log-link function;
3. the "classic APC" model of Hobcraft et al. (1982) and others;
4. the "reduced Plat" ("RP") model of Plat (2009a) discussed in Chapter 4;[18] and
5. the model produced by the "general procedure" ("GP") in Chapter 9 for the data described above.

These models have the forms

$$\ln(\mu_{x,t}) = \alpha_x^{(LC)} + \beta_x^{(LC)}\kappa_t^{(LC)} \tag{11.24}$$

$$\ln(\mu_{x,t}) = \alpha_x^{(CBDX)} + \kappa_t^{(CBDX,1)} + (x-\bar{x})\kappa_t^{(CBDX,2)} \tag{11.25}$$

$$\ln(\mu_{x,t}) = \alpha_x^{(APC)} + \kappa_t^{(APC)} + \gamma_{t-x}^{(APC)} \tag{11.26}$$

$$\ln(\mu_{x,t}) = \alpha_x^{(RP)} + \kappa_t^{(RP,1)} + (x-\bar{x})\kappa_t^{(RP,2)} + \gamma_{t-x}^{(RP)} \tag{11.27}$$

$$\ln(\mu_{x,t}) = \alpha_x^{(GP)} + \sum_{i=1}^{3} f^{(GP,i)}(x)\kappa_t^{(GP,i)} + \gamma_{t-x}^{(GP)} \tag{11.28}$$

where $f^{(GP,1)}(x) = 1$, $f^{(GP,2)}(x) = (x - x_2)^+$ and $f^{(GP,3)}(x) = (x_3 - x)^+$.[19] The parameters in these models have been estimated by fitting the model to the UK population data described above. These fitted parameters have, in turn, been used to estimate the

[18] That is, the simplification of the main model discussed in Plat (2009a) without the third, high-age term or, equivalently, an extension of the CBDX model with a cohort term.

[19] In this, the ages x_2 and x_3 in $f^{(GP,2)}(x)$ and $f^{(GP,3)}(x)$ are free parameters found by maximising the fit to data, which take the values $x_2 = 73$ and $x_3 = 84$ for the data in question. We also select age functions which are normalised so that $\sum_x |\beta_x| = \sum_x |f(x)| = 1$. This involves either including normalisation constants or choosing age functions which are "self-normalising" in the sense of Chapter 3. However, for clarity, these are not shown, although they are taken into account in the fitting algorithms.

parameters of the time series processes discussed in Sections 11.2.4.1 and 11.2.4.2 for $\boldsymbol{\kappa}_t$ and γ_y (if applicable). Using these parameter estimates, we can calculate forward mortality rate surfaces in the real-world measure using Equation 11.12.

These models have been chosen to give a reasonable cross section of the different APC mortality models which could be used in practice. It should be noted that for most of these models, we can use a constant drift function in Equation 11.14 and the projections will be well-identified. The exception to this is the reduced Plat model, where the random walk for $\kappa_t^{(RP,1)}$ requires a linear drift in order to be well-identified, as discussed in Chapter 4.

One of the advantages of the forward mortality rate framework described in this paper is that it allows for consistency between the model of the force of mortality and the forward mortality surface. Consequently, as a check, we compare these forward surfaces of mortality for each model to the mean mortality rates calculated using Monte Carlo simulations (shown in Figure 11.1 for the GP model) and find that the small difference between the two is explained by sampling error in the simulations.

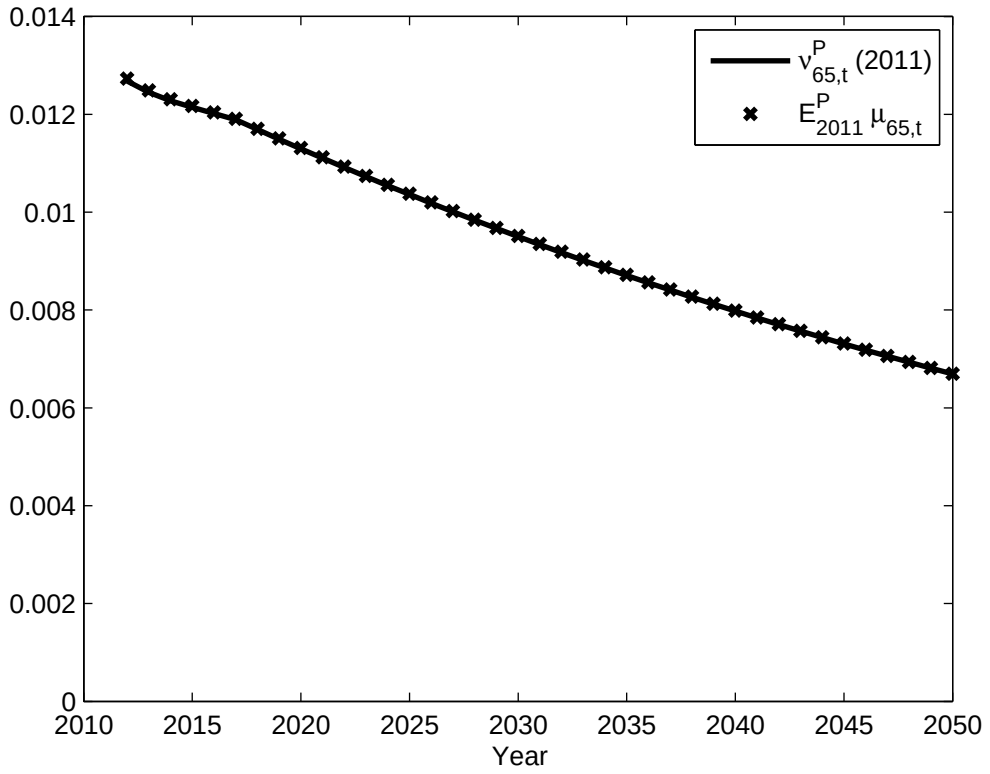


FIGURE 11.1: Difference between forward mortality rates and those obtained from Monte Carlo simulations using the GP model

11.3 Pricing securities and the market price of longevity risk

11.3.1 The market-consistent measure

In Section 11.2.4, we calculated mortality forward rates using the time series processes estimated from the fitted parameters. This means that the expectations in Equation 11.13 were calculated in the historical, real-world measure, $\mathbb{P}$.

It is obviously important that longevity-linked securities prices are consistent across different types of security in order to limit the potential for pricing anomalies and arbitrage opportunities in the market. In addition, modern solvency regimes require that liability values and technical provisions for pension schemes and insurers must also be consistent with market prices. Identifying a suitable market-consistent measure, $\mathbb{Q}$, is therefore a critical component of the forward mortality framework.

The starting point of modern financial theory is to assume that the financial markets are "complete" in the sense that every financial claim in them can be hedged perfectly using tradable assets. In complete markets, the market-consistent measure exists and is unique. Derivative securities in complete markets can be perfectly replicated using these underlying securities without risk (and hence these measures are also referred to as "risk-neutral") and the costs of these hedging strategies give the derivatives their unique prices. Complete markets are also free from arbitrage, since all prices can be derived using these underlying hedging strategies and any deviation from these prices will be arbitraged away by informed investors. The assumption of market completeness is a reasonable one in many contexts, such as developed markets for equities and interest rates in large and advanced economies.

However, the market for longevity risk is not complete. Not only are there insufficient tradable longevity-linked securities to fully replicate all financial claims, there are almost no longevity-linked securities being actively traded, full stop. Therefore, defining a market-consistent measure for longevity risk is a major problem for all mortality models which seek to price longevity-linked securities.

Some studies, for instance Schrager (2006), assume a priori that any market will be risk-neutral with respect to longevity risk and therefore that the historical and market-consistent measures are equal. We believe this is unlikely, given that any market in longevity risk is likely to be dominated by parties which suffer financially from rising life expectancy (see Loeys et al. (2007)) and therefore will be generally seeking to hedge the risk of future improvements in mortality rates.

In light of this absence of information, Barrieu et al. (2012, p. 224) suggested that the real-world measure must play a key role in the definition of any market-consistent measure:

> What will be a good pricing measure for longevity? It is expected that the historical probability measure will play a key role, due to the reliable data associated with it. Therefore, it seems natural to look for a pricing probability measure equivalent to the historical probability measure. Important factors to consider are that a relevant pricing measure must be: robust with respect to the statistical data, and also compatible with the prices of the liquid assets quoted in the market. Therefore, a relevant probability measure should make the link between the historical vision and the market vision. Once the subsets of all such probability measures that capture the desired information are specified, a search can commence for the optimal example by maximising the likelihood or the entropic criterion.

We agree with this analysis, and use the Esscher transform to define a market-consistent measure that is equivalent to the real-world measure and that satisfies many of these desirable properties. This transformation is relatively parsimonious, with a small number of free parameters which can be calibrated using any market information we possess. Below, we further show that the Esscher transform gives us closed form expressions for the market-consistent forward mortality rates as shown below, and therefore is relatively straightforward to implement and robust to calibrate to data.

The Esscher transform has often been used in securities pricing in imperfect markets since the work of Gerber and Shiu (1994). As discussed in Kijima (2005), it is related to other widely used distortion methods for adjusting to a risk-neutral measure, such as the the Wang transform (developed in Wang (2000, 2002) and Cox et al. (2006), and used in Denuit et al. (2007) for example), and the Sharpe ratio in modern financial theory (used in Milevsky et al. (2005) and Loeys et al. (2007)). It is also consistent with pricing in the real-world measure for an individual with an exponential utility function, as discussed in

Milidonis et al. (2011).

For a risk $X_{x,t}$ in the $\mathbb{P}$ measure, the general Esscher transform to the $\mathbb{Q}$ measure can be defined by

$$\mathbb{E}^{\mathbb{Q}} X_{x,t} = \frac{\mathbb{E}^{\mathbb{P}} \left[X_{x,t} \exp(-Z_{x,t}) \right]}{\mathbb{E}^{\mathbb{P}} \exp(-Z_{x,t})} \tag{11.29}$$

where $Z_{x,t}$ is a random variable containing the parameters defining the market-consistent measure.

In the context of mortality forward rates, we choose $X_{x,t} = \mu_{x,t} = \exp(\eta_{x,t})$ and correspondingly define

$$Z_{x,t} = \boldsymbol{\lambda}^{\top} \boldsymbol{\kappa}_t + \lambda^{\gamma} \gamma_{t-x} \tag{11.30}$$

where $\boldsymbol{\lambda}$ is an $(N \times 1)$ column vector. Hence, there are $N+1$ parameters (which we refer to collectively as $\lambda^{(j)}, j \in \{1, \dots N, \gamma\}$), which correspond to the N age/period terms (in the vector $\boldsymbol{\lambda}$), and the cohort term (with single parameter $\lambda^{(\gamma)}$) in the general APC mortality model in Equation 11.2. It is important to note that the values found for these parameters will depend upon the specifics of the underlying model, and so are not comparable between different models.

Due to the paucity of genuine market information to price longevity risk, one might have a natural inclination to prefer simpler models, such as the LC model (which has only one free parameter for the Esscher transform). Such models could be felt to be more parsimonious, having fewer market prices for longevity risk and therefore requiring fewer market prices for longevity-linked securities in order to calibrate the market-consistent measure. For example, calibrating the LC model would require only one market price in order to calibrate the market-consistent measure, whilst calibrating the GP model in Section 11.2.5 requires four market prices. Using overly simple models, however, would be a mistake which can lead to unreasonable prices for other longevity-linked securities as shown in Section 11.3.3.

Using the Esscher transform with Equation 11.11 and this definition for $Z_{x,t}$ gives

$$\begin{aligned}
\nu^{\mathbb{Q}}_{x,t}(\tau) &= \mathbb{E}^{\mathbb{Q}}{}_{\tau}\ \mu_{x,t} \\
&= \mathbb{E}^{\mathbb{Q}}{}_{\tau}\ \exp(\eta_{x,t}) \\
&= \frac{\mathbb{E}^{\mathbb{P}}{}_{\tau}\ \exp(-Z_{x,t}\eta_{x,t})}{\mathbb{E}^{\mathbb{P}}{}_{\tau}\ \exp(-Z_{x,t})} \\
&= \frac{\mathbb{E}^{\mathbb{P}}{}_{\tau}\ \exp(\alpha_x + (\boldsymbol{\beta}_x - \boldsymbol{\lambda})^{\top}\boldsymbol{\kappa}_t + (1-\lambda^{\gamma})\gamma_{t-x})}{\mathbb{E}^{\mathbb{P}}{}_{\tau}\ \exp(-\boldsymbol{\lambda}^{\top}\boldsymbol{\kappa}_t - \lambda^{\gamma}\gamma_{t-x})} \\
&= \exp\left(\alpha_x + \boldsymbol{\beta}_x^{\top}\mathbb{E}^{\mathbb{P}}{}_{\tau}\ \boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x^{\top}\mathbb{V}ar^{\mathbb{P}}_{\tau}\ (\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{E}^{\mathbb{P}}{}_{\tau}\ \gamma_{t-x}\right. \\
&\quad \left. + \frac{1}{2}\mathbb{V}ar^{\mathbb{P}}_{\tau}\ (\gamma_{t-x}) - \frac{1}{2}\boldsymbol{\beta}_x^{\top}\mathbb{V}ar^{\mathbb{P}}_{\tau}\ (\boldsymbol{\kappa}_t)\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{\lambda}^{\top}\mathbb{V}ar^{\mathbb{P}}_{\tau}\ (\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x - \lambda^{\gamma}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\gamma_{t-x})\right) \\
&= \exp\left(-\boldsymbol{\beta}_x^{\top}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} - \lambda^{\gamma}\mathbb{V}ar^{\mathbb{P}}_{\tau}(\gamma_{t-x})\right)\nu^{\mathbb{P}}_{x,t}(\tau) \qquad (11.31)
\end{aligned}$$

due to the symmetry of $\mathbb{V}ar^{\mathbb{P}}_{\tau}(\boldsymbol{\kappa}_t)$.

This gives us closed-form expressions which allow us to adjust the forward mortality rates in the real-world measure to a market-consistent measure. The existence of closed-form expressions is why we argued that the Esscher transform neatly complements the forward mortality framework: these results could not have been achieved with alternative transformations to the market-consistent measure. Since we have already found expressions for $\mathbb{V}ar^{\mathbb{P}}_{\tau}(\boldsymbol{\kappa}_t)$ and $\mathbb{V}ar^{\mathbb{P}}_{\tau}(\gamma_y)$, transforming the forward mortality surface in the real-world measure into a market-consistent measure is simply a matter of finding the values of free parameters of the Esscher transform. This can be done if we have sufficient prices for longevity-linked securities, as discussed in Section 11.3.2 below.

Through the analogy with utility pricing and the Sharpe ratio, we refer to the parameters of the Esscher transform as the "market prices of longevity risk" associated with each of the age/period and cohort terms. For this analogy to be reasonable, we would anticipate that the parameters, $\lambda^{(j)}$, should be positive. However, this is not necessarily the case in the forward mortality framework, for the following reasons.

As discussed in Loeys et al. (2007), we anticipate that the marginal participant in the market for longevity-linked securities will be a life insurer seeking to hedge longevity risk. Such a life insurer will be averse to longevity risk, and so, we would expect the market-consistent forward mortality rates to be lower than those in the real-world measure

$$\nu^{\mathbb{Q}}_{x,t}(\tau) \leq \nu^{\mathbb{P}}_{x,t}(\tau)$$

In order for this to be true,

$$\begin{aligned}\exp\left(-\boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau^{\mathbb{P}}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} - \lambda^\gamma \mathbb{V}ar_\tau^{\mathbb{P}}(\gamma_{t-x})\right) &\leq 1 \\ \Rightarrow \boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau^{\mathbb{P}}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar_\tau^{\mathbb{P}}(\gamma_{t-x}) &\geq 0\end{aligned}$$

Since $\mathbb{V}ar_\tau^{\mathbb{P}}(\boldsymbol{\kappa}_t)$ is a positive definite matrix and $\mathbb{V}ar_\tau^{\mathbb{P}}(\gamma_y) \geq 0$, this will certainly be true if $\lambda^\gamma > 0$ and the elements of $\boldsymbol{\lambda}$ are also positive. However, individual market prices of longevity risk can be negative, whilst still ensuring that hedgers pay a positive price to transfer longevity risk overall. Since some market prices can be negative, the term "market prices" might be considered misleading. Although we shall refer to these parameters as market prices in this chapter and in Chapter 12, it should be borne in mind that they are probably best thought of as simply parameters in the Esscher transform in Equation 11.29 rather than true market prices of longevity risk based on an expected utility approach (such as that discussed in Zhou et al. (2015)).

The Esscher transform approach has some other practical advantages, beyond the existence of closed-form expressions for the forward mortality rates. The forward mortality surface in the real-world measure will be updated only infrequently, typically once every year when new mortality data is released. However, market information will need to be updated far more frequently, especially as the market for longevity-linked securities develops. It is desirable in practice to be able to take the (infrequently changing) $\mathbb{P}$-measure forward mortality surface and make relatively simple adjustments to this to reflect changing market information, rather than having to re-estimate the model completely every time the pricing information changes.

However, a limitation of the forward mortality framework outlined in this study is that it is currently unable to price longevity-linked securities with optionality, for example, a call option on mortality rates. In order to do this, the dynamics of mortality rates in the market-consistent measure would need to be specified, in addition to simply the expectation, $\mathbb{E}^{\mathbb{Q}}{}_\tau\, \mu_{x,t}$. We leave the extension of the forward mortality framework to the inclusion of longevity-linked options to future work.

We also note that, looking solely at the age/period terms, Equations 11.16 and 11.17 imply

$$\begin{aligned}\boldsymbol{\beta}_x^\top \mathbb{E}^{\mathbb{P}}{}_\tau \boldsymbol{\kappa}_t + \boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} &= \boldsymbol{\beta}_x^\top \left[\boldsymbol{\kappa}_\tau + \mu\chi_{\tau,t} + (t-\tau)\Sigma\boldsymbol{\lambda}\right] \\ &= \boldsymbol{\beta}_x^\top \left[\boldsymbol{\kappa}_\tau + \hat{\mu}\chi_{\tau,t}\right]\end{aligned}$$

since $t - \tau$ is always one of the deterministic functions in $\chi_{\tau,t}$. Hence, we see that for an age/period model such as the LC and CBDX models, the Esscher transform to the market-consistent measure is equivalent to making an adjustment to the drift of the random walk in Equation 11.14. In this form, the use of the Esscher transform can be compared with some of the other approaches that have been suggested in previous studies. For instance, Loeys et al. (2007) suggested that the price of a q-forward should be calculated as

$$q^f = (1 - (t - \tau)\tilde{\lambda}\sigma^2)q^e$$

where σ^2 is defined as the annual volatility of the mortality rate, i.e., $\sigma^2 = \mathbb{V}ar^{\mathbb{P}}(\ln q)$. We can compare this pricing formula to what our forward mortality framework would give were we to use the LC model as the underlying mortality model. This has one period function, κ_t, with one associated market price of risk, λ. From Equation 11.31 applied to the LC model, we find

$$\nu^{\mathbb{Q}}_{x,t}(\tau) = \exp\left(-(t - \tau)\beta_x \Sigma\lambda\right) \nu^{\mathbb{P}}_{x,t}(\tau)$$

We can therefore see that the pricing formula in Loeys et al. (2007) is similar in form to Equation 11.31, although based on forward contracts on probabilities of death, $q_{x,t}$, rather than the longevity-zeros which are used as the underlying securities in this study.

Cairns et al. (2006a) adjusted the drift of the random walk used to project the period functions directly, in order to incorporate market prices for longevity risk without recourse to the Esscher transform

$$\mu^{\mathbb{Q}} = \mu^{\mathbb{P}} - C\tilde{\boldsymbol{\lambda}}$$

where $CC^\top = \Sigma$ and $\boldsymbol{\lambda}$ is a vector of the market prices of risk. If such an approach were to be used for the CBDX model in a forward mortality rates framework such as above, we would find market-consistent forward mortality rates

$$\nu^{\mathbb{Q}}_{x,t}(\tau) = \exp\left(-(t - \tau)\boldsymbol{\beta}_x^\top C\tilde{\boldsymbol{\lambda}}\right) \nu^{\mathbb{P}}_{x,t}(\tau)$$

Therefore, we see that the approach used in Cairns et al. (2006a) is equivalent to that used in this study, except using $C\tilde{\boldsymbol{\lambda}}$ instead of $\Sigma\boldsymbol{\lambda}$. Equating these gives

$$C\tilde{\boldsymbol{\lambda}} = \Sigma\boldsymbol{\lambda}$$
$$\tilde{\boldsymbol{\lambda}} = C^\top\boldsymbol{\lambda}$$

Hence, the more rigorous forward mortality framework defined in this study achieves results which are consistent with those of Cairns et al. (2006a), but is also able to justify the otherwise ad hoc adjustments to the drift made in that study.

11.3.2 Calibration of the market-consistent measure

As has been mentioned previously, a major problem with forward mortality models is the lack of market information to specify the market-consistent measure. An advantage of using the forward mortality framework described in this study is that, rather than requiring sufficient market prices to define the full forward mortality surface, we require only $N + 1$ prices to uniquely specify the market prices of longevity risk used in the Esscher transform. This substantially reduces the market information required.

However, even this simplification is unlikely to be adequate at present, given the paucity of traded longevity-linked securities. Many of those which do exist, such as the extreme mortality bonds listed in Lane (2011), are not suitable as they involve options on mortality rates which cannot be priced using the forward mortality framework in its current state of development. For illustrative purposes, we will demonstrate how the forward mortality rate framework could be calibrated with respect to the sort of information which is available currently or is likely to be available in the foreseeable future, and how this "external" market in longevity risk could be supplemented by use of an "internal" market for longevity risk based on the assumptions used to value and reserve for longevity risk within a life insurer.[20]

11.3.2.1 External market

A number of "external" markets exist for products which depend upon longevity, for instance the markets for endowment assurances and individual annuities. These were used to provide market information for pricing longevity risk in Bayraktar and Young (2007) and Bauer et al. (2008). However, both of these products are sold to individuals, and therefore are subject to idiosyncratic mortality risk as well as systematic longevity risk, which makes them unsuitable for use in a forward mortality rate framework, as discussed by Norberg (2010). Furthermore, insurers will include loadings for expenses and other risks, in addition to longevity risk when pricing these products, which makes using them

[20] In a sense, the difference between the external and internal markets for longevity risk could be compared to the difference between using mark-to-market and mark-to-model valuation methods when valuing securities in company accounts, depending upon whether deep and liquid markets exist for them.

to calibrate a forward mortality model problematic.

Instead, any forward mortality model will need to be calibrated using securities dependent on aggregate mortality rates (preferably from national populations) rather than those that are sold to individuals. Such securities are also more likely to be traded, thereby giving informed and responsive market prices. The problem remains, however, that there is currently no actively-traded market in such securities which can be used to provide the pricing information required to calibrate the market-consistent measure.

To date, probably the most active market in longevity-linked securities has been that for bespoke longevity swaps (see Chapter 10). A longevity swap is an agreement between two parties to swap a series of cashflows - a fixed leg based on the best estimate of the survivorship of a cohort but then increased by a constant percentage (the swap margin) and a floating leg based on the actual survivorship observed for the cohort. A bespoke longevity swap is one which is tailored to the characteristics of a specific population such as a pension scheme. As such, bespoke longevity swaps are unlikely to be widely traded, and act more as customised reinsurance contracts than standardised longevity-linked securities which could form the basis for a market in longevity risk. In contrast, an index-based swap, such as that described in Dowd et al. (2006b), is one where the cohort in question is from a national population. Although index-based longevity swaps have not yet been widely traded, the development of the bespoke longevity swap market to date implies that, if a market in longevity risk does develop in the near future, it is likely that index-based swaps will form a key component of it.

For illustrative purposes, we therefore assume the existence of a single index-based longevity swap, which we believe might be typical of the sort of security which may be traded during the early stages of the development of an external market in longevity-linked securities. We assume that this index-based longevity swap has been written on a standard cohort of men in the UK aged 65 in 2011 and has a term of 35 years (i.e., until the cohort is aged 100). The floating leg of this swap will therefore have the value

$$\sum_{t=1}^{35} {}_tP^{\mathbb{Q}}_{65,\tau}(\tau)B(\tau,\tau+t)$$

i.e., the same price as a series of the longevity zeros discussed in Section 11.2.2. The fixed-leg cashflows will reflect a typical "best estimate" agreed between the contracting parties when the swap is initiated. For illustrative purposes, we assume these cashflows are set by calculating the survivorship of the reference cohort using the fitted mortality

rates in $\tau = 2011$ projected using the "CMI Projection Model" (Continuous Mortality Investigation (2009a,b, 2013)) with a "long-term rate of improvement" assumption of 1.5% p.a..[21] We denote the survival probabilities of the reference cohort from time τ to $\tau + t$ using this assumption as ${}_t\tilde{P}_{65,\tau}(\tau)$. While there is currently no active market in index-based swaps, this assumption is typical of those used to define the fixed leg of bespoke longevity swaps in our experience. These cashflows are then increased by a swap premium of 4%, which is a typical level on bespoke swaps in our experience.

The price of the swap is therefore

$$\sum_{t=1}^{35} \left({}_tP^{\mathbb{Q}}_{65,\tau}(\tau) - 1.04\ {}_t\tilde{P}_{65,\tau}(\tau) \right) B(\tau, \tau + t) \tag{11.32}$$

and will be zero at time τ. We therefore calibrate the market prices of risk to impose this using standard numerical optimisation algorithms. In these calculations, we assume a flat real yield of 1.0% p.a. for the zero-coupon bond prices, $B(\tau, \tau + t)$

For models with only one source of risk (for instance, the LC model), this single, external price is sufficient to specify the single market price of longevity risk uniquely. For more complicated models, with multiple risk sources, we require additional prices in order to specify the market prices of longevity risk.

11.3.2.2 Internal market

We observe that, while genuine market information is in scarce supply, many insurance companies will effectively have an internal market for longevity risk due to the cross-subsidies between different lines of business with different exposures to longevity risk. For instance, an insurer which writes both annuity and life assurance lines of business has, de facto, established an internal market for longevity risk due to the presence of natural hedging between the two lines of business, as discussed in Cox and Lin (2007). The "price" of longevity risk in this internal market will find expression in the mortality improvement assumptions used in the pricing and reserving for these different lines of business. It is therefore natural to use these "internal" market signals to supplement

[21] The use of the CMI Projection Model in this context is purely illustrative and should not imply that we believe that this is the best model to use for pricing longevity-linked securities, although it is typical of what has been used in practice in our experience.

those coming from the genuine external market if there are insufficient traded longevity-linked securities to define the market-consistent measure.

Alternatively, an insurer may develop an "internal" price for longevity risk by analysing the cost of longevity reinsurance via bespoke longevity swaps. Although these contracts do not solely transfer longevity risk, since they also transfer basis and idiosyncratic risks, they could still give some indication of a price for the systematic longevity risk present, and so be used to calibrate the market-consistent measure.

For example, we assume that the forward mortality framework is being used by an organisation with an internal, deterministic assumption that constitutes their "house view" of mortality improvements. This house view would then feed through into the assumptions used in pricing and reserving, and inform those assumptions that are used for accounting and regulatory purposes if there is sufficient flexibility in how these are set. The existence of such a house view would therefore determine the organisation's appetite for longevity risk across multiple lines of business and so underpin the "internal" market for longevity risk.

To illustrate the sort of internal market that might be considered typical, we assume a house view that mortality rates improve in line with the projections from the CMI Projection Model with a long-term rate of improvement of 1.75%.[22] Again, this is in line with the sort of assumptions used to reserve for and price annuity business in the UK in our experience. In order to translate this house view into the market prices of longevity risk in our forward mortality framework, we try to minimise the (weighted) relative distance between the surface of probabilities of dying given by the internal assumption, $\tilde{q}_{x,t}$, and those given the forward mortality surface in the $\mathbb{Q}$-measure

$$\mathcal{Q}_{x,t}(\tau) = 1 - \exp\left(-\nu^{\mathbb{Q}}_{x,t}(\tau)\right)$$

[22]This value of 1.75% can be compared with the assumption of a long-term rate of improvement of 1.5% used for the fixed leg of the index-based longevity swap above. The long term rate of improvement is likely to be higher on an annuity reserving basis than for valuing a longevity swap, since it is common practice, in our experiences, for annuity providers to include an implicit margin for prudence in their mortality projection. In contrast, the assumption used in a longevity swap typically reflects a best estimate of future mortality improvements and risk is explicitly allow for via the swap premium rather than an implicit margin in the mortality assumption.

at certain key ages, subject to the swap also being priced fairly at time, τ, i.e.,

$$\min_{\lambda} \sum_{t,x \in \mathcal{X}} B(\tau, \tau+t) \frac{\left(\tilde{q}_{x,t} - \mathcal{Q}_{x,t}\right)^2}{\tilde{q}_{x,t}}$$
$$\text{subject to Equation 11.32 } = 0$$

where $\mathcal{X} = \{50, 55, 60, 65, 70, 75, 80\}$. This procedure is equivalent to determining the market-consistent measure by reference to an external market in q-forwards, as proposed in Coughlan et al. (2007b) and discussed in Section 11.3.3.2 below, if such as market existed. We consider these key ages partly to ensure that the forward mortality surface in the market-consistent measure is biologically reasonable over a wide age range and because, if a market in q-forwards does emerge, it is at these ages where the market is likely to be most liquid (see Li and Luo (2012)). Therefore, the use of the internal market for longevity risk is simply a proxy for information from an external market for longevity risk, and will be supplanted should a genuine external market develop.

We use these assumptions for the external and internal markets for longevity risk in order to calibrate the parameters of the Esscher transform for all five models described in Section 11.2.5. These parameters, along with the forward mortality surfaces obtained in Section 11.2.5, allow us to construct the forward mortality surface in the market-consistent measure, which can then be used to value other longevity-linked liabilities and securities in a market-consistent fashion.

11.3.3 Pricing longevity-linked securities

The forward mortality framework described above provides a single surface of forward mortality rates, calibrated from all the available information on longevity-linked securities. It can, therefore, be used to value any other longevity-linked securities and give prices consistent with those observed. We demonstrate this for a range of different longevity-linked securities below.

11.3.3.1 Survivor derivatives

Longevity zeros and s-forwards

In Section 11.2.2, we defined the forward mortality rates assuming the existence of a market in longevity zeros. These were used as they are the fundamental securities dependent upon the survivorship of a cohort of individuals, and can be used to construct

more complicated survivor securities such as annuities and longevity swaps, as discussed below. Related to longevity zeros are "s-forwards", as proposed in Dowd (2003), Blake et al. (2006) and the Life and Longevity Markets Association,[23] which are forward contracts defined on a longevity zero (and hence are more capital efficient).

From Equation 11.7, we can see that

$$\mathcal{S}_{x,t}(\tau) = {}_tP^{\mathbb{Q}}_{x,\tau} = \exp\left(-\sum_{u=1}^{t} \nu^{\mathbb{Q}}_{x+u,\tau+u}(\tau)\right)$$

where $\mathcal{S}_{x,t}(\tau)$ is the forward price of an s-forward at time τ, defined on a cohort aged x at τ, with a maturity of t years. Figure 11.2 shows s-forward prices defined on the cohort of individuals aged 65 in 2011 with different maturities.

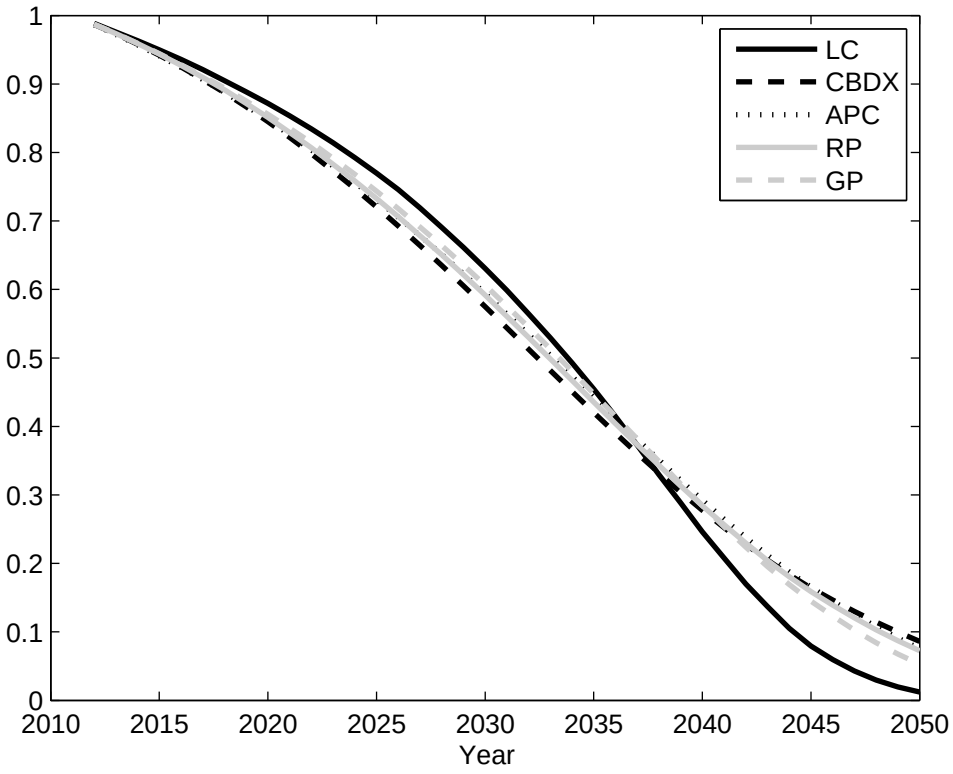


FIGURE 11.2: S-forward prices for five different mortality models

As can be seen, most of the models give broadly comparable s-forward prices, especially those calibrated using the internal market information. We note that the LC model gives s-forward prices which are slightly different from these models, with higher probabilities

[23] http://www.llma.org/

of survival over the first few decades followed by a period of higher mortality rates (and hence a steeper gradient for the curve), but these are still biologically reasonable.

Annuities

The most relevant longevity-linked instruments for many life insurance companies are annuities. For the reasons discussed in Section 11.3.1 and Norberg (2010), individual annuities cannot be used to calibrate the forward mortality surface in the market-consistent measure, since the cashflows of these instruments are explicitly linked to the survivorship of a named individual and, hence, their prices include an allowance for individual mortality risk. In addition, they are not traded, and, therefore, cannot provide timely information on their values. However, when a life insurer reserves for a book of annuities, the idiosyncratic mortality risks are diversifiable and so are not included in the value of any specific annuity but through the additional capital required for the book.[24] In addition, modern solvency regimes, such as Solvency II, require the best estimate of the liabilities in respect of annuity policies to be calculated using market-consistent assumptions. Therefore, the market-consistent forward framework could, potentially, be used as the basis for an insurer's "internal model" under Solvency II, as discussed in EIOPA (2014).[25]

The value of an annuity can be directly constructed from a portfolio of longevity zeros using

$$a_x(\tau) = \sum_{t=0}^{\infty} {}_tP^{\mathbb{Q}}_{x,\tau}(\tau)B(\tau,\tau+t) \tag{11.33}$$

To calculate the values of longevity zeros beyond the maximum age in our data, we use the topping out procedure of Denuit and Goderniaux (2005). We therefore see that annuity values are very closely related to the swap price given in Equation 11.32.We calculate annuity prices[26] for men at different ages in 2011 using the five different models, and the results are shown in Figure 11.3.

[24] There will therefore be a distinction between the price an annuity is sold to the public for and the amount it is reserved for by the life insurer, with the additional margin for idiosyncratic mortality risk charged to the individual forming part of the profit margin of the product.

[25] This is discussed further in Chapter 12.

[26] Annuities are valued using a real discount rate of 1% p.a..

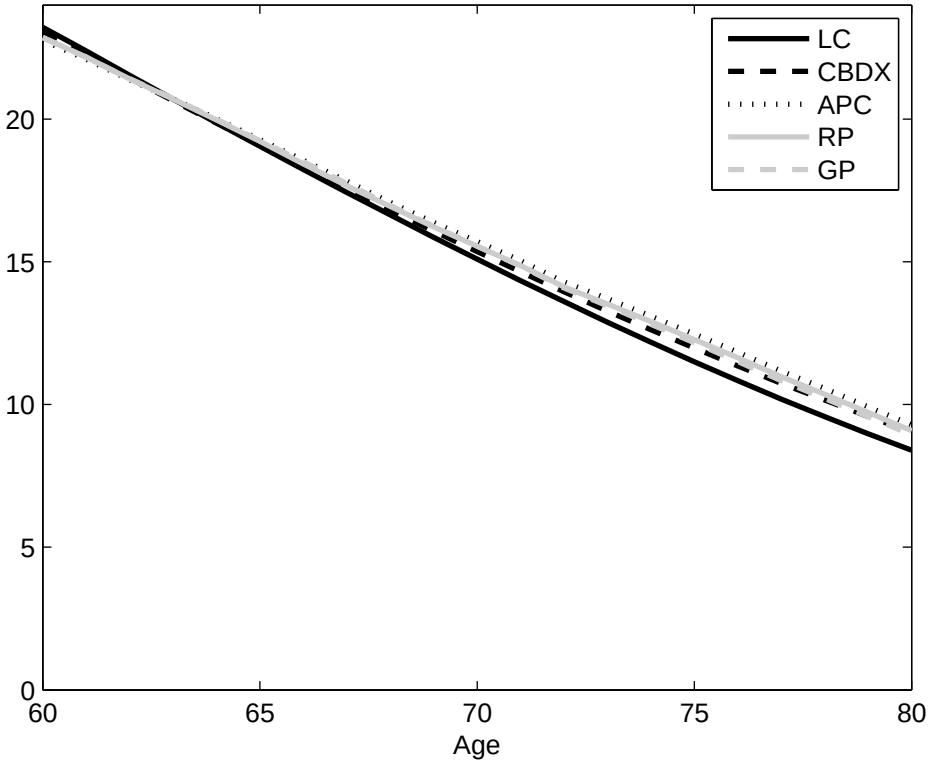


FIGURE 11.3: Annuity values for five different mortality models

We can see from this that the different models give broadly similar annuity values. This is not surprising given that they all use the same external market information (i.e., the swap price) in order to calibrate the market-consistent measure. Indeed, all the models give exactly the same value for an annuity at age 65, since this is determined by the swap price we have assumed and an annuity is equivalent to the floating leg of a longevity swap. However, the annuity values given by different models diverge slightly as we move away from this fixed reference point, with the LC model giving lower annuity values at higher ages than the other models.

Index-based longevity swaps

We can also use these results to investigate the potential pricing of index-based longevity swaps at different ages. Extending the definition of the swap value in Equation 11.32 for different ages to

$$0 = \sum_{t=1}^{35} \left({}_tP^{\mathbb{Q}}_{x,\tau}(\tau) - (1+\pi)\ {}_t\tilde{P}_{x,\tau}(\tau) \right) B(\tau, \tau+t) \tag{11.34}$$

we can use the same "best estimate" assumption based on the CMI Projection Model for the fixed legs of the swaps, to calculate the implied swap premium, π, on index-based longevity swaps at different ages. The implied swap premiums are shown in Figure 11.4.

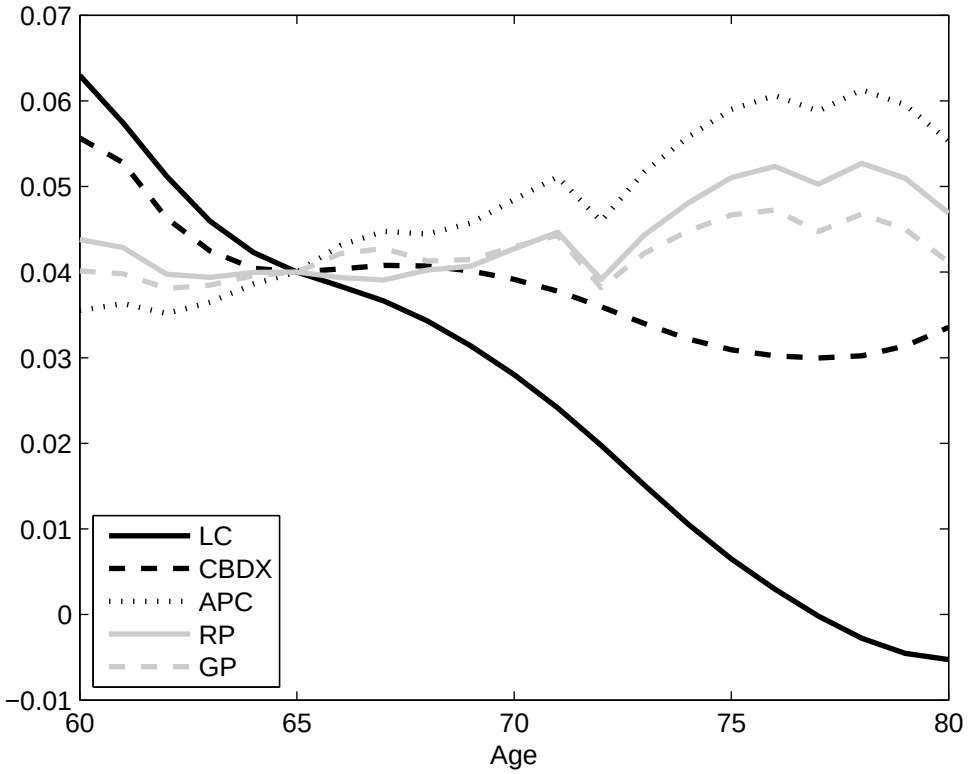


FIGURE 11.4: Swap premiums for five different mortality models

As can be seen, the behaviour of the swap premium depends strongly upon the model being used. For the classic APC, RP and GP models, which include a cohort term, the swap premium slightly increases with age, from around 3% at age 60 to around 6% between ages 75 and 80 (note that a value of 4% was assumed at age 65). Swap premiums in the CBDX model are relatively high at the youngest ages (5.5% at age 60) and decrease slowly with age, to around 3% at age 75. However, for all of these models, the swap premium remains positive and do not appear unreasonable at any age.

In contrast, the LC model gives swap premiums which decrease rapidly with age, giving negative swap premiums at higher ages (i.e., a premium would be paid to receive the floating payments on the swap) which does not appear reasonable. This is because the LC model gives relatively low values for annuities at higher ages - lower than would be found using the deterministic CMI Projection Model. We therefore see that there is a trade-off. On the one hand, we would like to use simple models which have relatively

few free parameters and so are simple to calibrate from sparse data (and, in particular, would avoid the use of an internal market for longevity risk). On the other hand, we also need to obtain plausible prices for different longevity-linked liabilities and securities and across a wide range of ages.

11.3.3.2 Other longevity-linked securities

A number of other longevity-derivatives not based on the survivorship of a cohort have been proposed, and these can also be valued using the forward mortality framework proposed here. A number of these are illustrated below. However, the important point to note is that any security which does not have a non-linear payoff (i.e., which is not an option) can be valued using the forward mortality framework proposed in this study.

q-forwards

Forward contracts on future probabilities of death, known as "q-forwards", were introduced in Coughlan et al. (2007b) represent another, distinct, family of potential longevity-linked securities. There have been a number of hedging transactions using q-forwards, as discussed in Blake et al. (2013), and so q-forwards are one of the major contenders to form the basis of a traded market for longevity risk if it develops. In addition, the internal market assumption, used in Section 11.3.2 to calibrate all of the models other than the LC model, implicitly makes use of a market for q-forwards, albeit one that is internal to the life insurer rather than an externally traded market.

Values for q-forwards at age 75 and different maturities, calculated using the forward mortality models, are shown in Figure 11.5, along with the $q_{x,t}$ values projected using the CMI Projection Model. For the models which used the internal market assumption to calibrate the market-consistent measure, we see that the q-forward values are broadly consistent with those from the CMI Projection Model. However, they are not identical, since the calibration process also has to match the swap price exactly and minimise the difference in q-forward prices at ages other than 75. However, because the GP model has more market prices of risk to calibrate, it achieves a slightly closer fit to the internal market assumption than the other models, including the cohort effect observed around 2025 (i.e., for cohorts born around 1950).

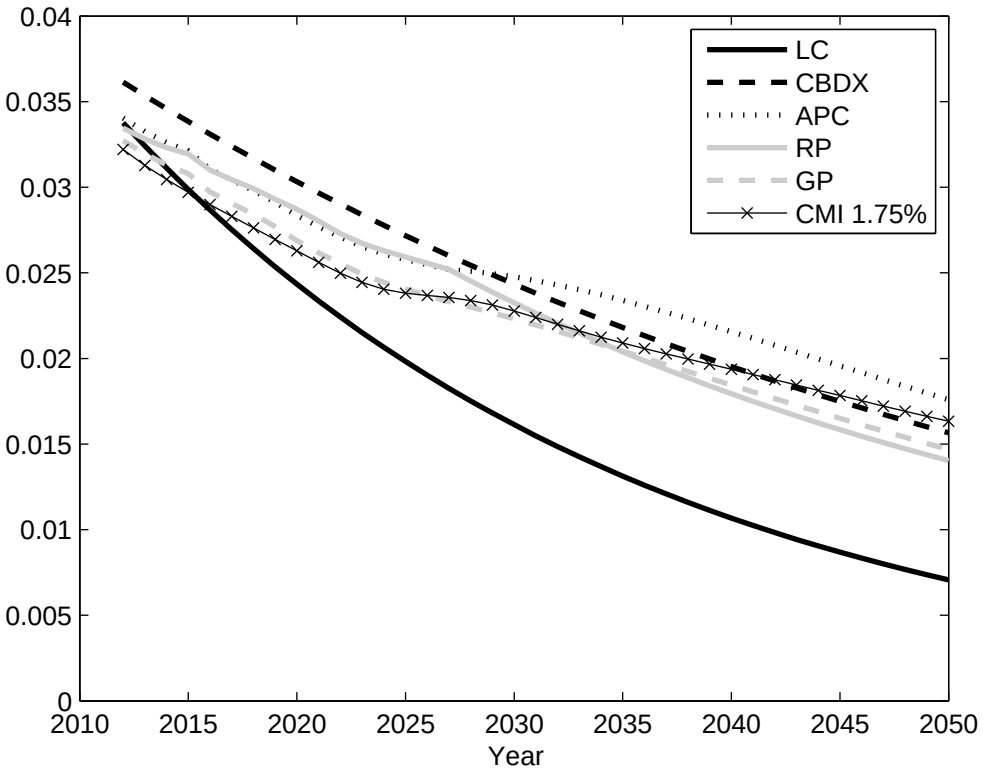


FIGURE 11.5: q-forward prices at age 75 for five different mortality models

In contrast, the LC model gives q-forward values which are very different from those of the other models, with implausibly rapid decreases in q-forward values. Again, this is because, with a single market price for longevity risk, the LC model has to severely distort the forward mortality surface in the real-world $\mathbb{P}$-measure in order to price the longevity swap. It cannot ensure that mortality rates across a wide range of other ages and years behave in a plausible fashion in the market-consistent measure. We therefore see that more sophisticated underlying APC mortality models, as well as being able to incorporate pricing information from a wider range of sources, will also tend to give more biologically-reasonable forward surfaces for mortality in the market-consistent measure.

e-forwards

Period life expectancy is a very commonly used aggregate measure of mortality rates, since it can be calculated easily from observed data and can be compared across different populations. It is, therefore, natural to consider its use as an index for longevity risk transfer, based on the suggestion of Denuit (2009). In particular, we consider a market

in forwards on period life expectancy, which we refer to as "e-forwards" (from the demographic symbol for period life expectancy). Using the forward mortality framework, we calculate forward period life expectancies as

$$\mathcal{E}_{65,t}(\tau) = 0.5 + \sum_{u=1}^{\infty} \exp\left(- \sum_{v=1}^{u} \nu^{\mathbb{Q}}_{65+v,t}(\tau) \right)$$

Figure 11.6 shows the forward period life expectancies at age 65 from each of the five models in the market-consistent measure.

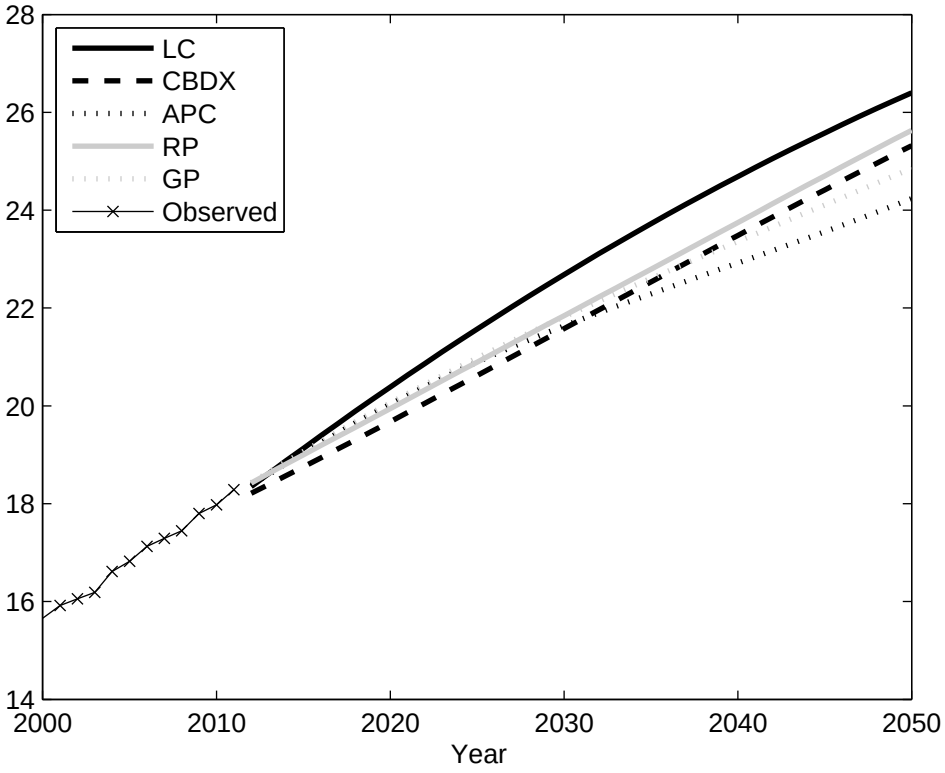


FIGURE 11.6: Period life expectancies at age 65 for five different mortality models

We note that all of the models give forward period life expectancies which can be considered biologically reasonable and consistent with the findings of Oeppen and Vaupel (2002), i.e., that they increase roughly linearly. Life expectancies from the LC model increase slightly faster than the other models, which otherwise give broadly consistent forward values. This is because of the use of the internal market to calibrate these other models, ensuring greater consistency between their forward mortality surfaces.

k-forwards

In Chapter 8, we discussed how the indices based on the observed rates of improvement in mortality rates, such as the indices which were defined in the construction of the Swiss Re Kortis bond, could potentially form the basis for a market in longevity risk. Improvement rates may be a natural basis for a market in longevity, as they are often used by actuaries to express long term assumptions regarding the evolution of mortality rates. Building on this, we also consider the forward value of the index for men in the UK defined by

$$\mathcal{K}_t(\tau) = \frac{1}{11}\sum_{x=75}^{85}\left(1 - \left[\frac{\nu^{\mathbb{Q}}_{x,t}(\tau)}{\nu^{\mathbb{Q}}_{x,t-8}(\tau)}\right]^{\frac{1}{8}}\right)$$

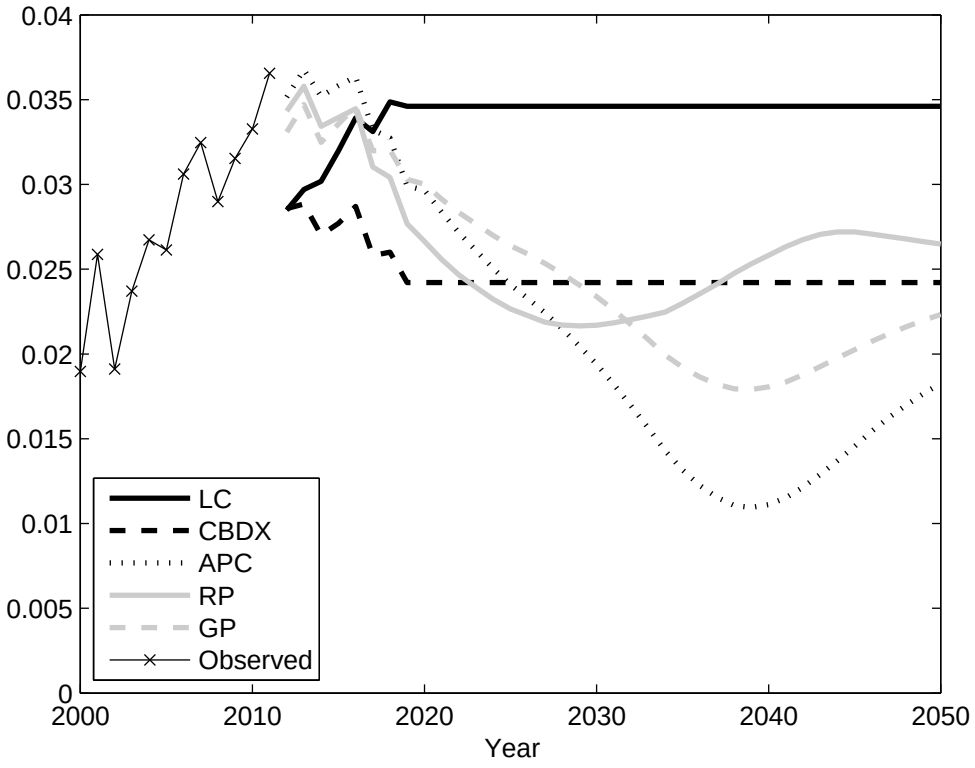
This index was constructed to measure the average rate of improvement in mortality rates between ages 75 and 85 for men in the UK and so could be used for hedging or transferring longevity risk in a portfolio of annuities. Unlike the Kortis bond, however, we only consider an index constructed for a single population (i.e., men in the UK) rather than the difference between two populations, and only consider pricing the index rather than an option on the index.[27]

In Chapter 8 it was suggested that forward contracts based on this Kortis index could form the basis of a market in longevity risk. We refer to such contracts as "k-forwards" in the same manner at q-, s- and e-forwards discussed above. Figure 11.7 shows the projected k-forward values in the market-consistent measure. As discussed in Chapter 8, the Kortis index is designed to be very sensitive to the rates of improvement in longevity, which are determined by the drift, μ, of the random walk used for the period parameters. Indeed, for models which lack a cohort term, the drift in the random walk exactly determines the projected index values, and hence they are constant beyond 2020.[28] For the models which include cohort parameters, the value of the index in the short term depends strongly upon the cohort parameters fitted by the model, as discussed in Chapter 8, resulting in a distinctive curved pattern. In general, the models containing a cohort term give market-consistent assumptions for the rate of improvement in longevity which decrease from its currently observed level of around 3.5% to around 2% in 20 years' time. This is not surprising given this is broadly in line with the assumptions used to calibrate the market-consistent measure, i.e., the CMI Mortality Projection Model with a long term rate of improvement of either 1.5% or 1.75%.

[27] See Chapter 8 for a further discussion of the Swiss Re Kortis bond and its construction.

[28] Before 2020, the Kortis index is based partly on projected and partly on observed mortality rates, and hence exhibits more variability than after 2020.

FIGURE 11.7: Kortis index values for five different mortality models



As in the case of the q-forwards, the index values for the LC model show a very different evolution due to the limited ability of this model to both price the market information and give a biologically reasonable forward surface of mortality. However, the alternative models appear to give index values which are biologically reasonable and consistent with the historical, realised values for the k-forwards, which potentially means that forwards on the index could form a viable basis for a market in longevity risk.

Other longevity-linked securities

The forward mortality surface could also be used to value life assurance policies in the same manner. In conjunction with the results of Chapter 12, the forward mortality framework could therefore be used as a standard model for both the valuation of a life insurer's technical provisions and the assessment of longevity risk within them, in accordance with the Solvency II regulatory regime described in EIOPA (2014). In addition, for life insurers writing both annuity and assurance policies, it may be desirable to value

these consistently in the technical provisions, in order to achieve the benefits from natural hedging discussed in Cox and Lin (2007).

Beyond the examples discussed above, the forward mortality framework could be used to value any longevity-linked security with a linear payoff in the underlying index. Hence, although the market for longevity-linked securities is in the early stage of development currently and it is unclear which form of securities will ultimately come to be traded, we believe that the framework described in this study is flexible enough to be able to price any of them in a manner consistent with any other prices for longevity-linked liabilities and securities which are available.

As discussed previously, one disadvantage of any forward mortality rate framework is that it cannot currently be used to value longevity-linked options, since it only looks at the expected mortality rates in the market-consistent measure. For example, it could not be used directly to value mortality catastrophe bonds, such as the Swiss Re Vita bond (discussed in Bauer and Kramer (2007)), Longevity Experience Options (described in Fetiveau and Jia (2014)), bespoke index-based solutions (described in Michaelson and Mulholland (2014)), a guaranteed annuity option (discussed in Pelsser (2003) and Ballotta and Haberman (2006)) or a bond similar to the Kortis bond with the principal being a non-linear function of the index value. At the present time, we do not think that this is a fatal limitation of the forward mortality rate framework discussed here, as currently the market for longevity-linked securities is not sufficiently developed to allow a full calibration of the forward mortality rate surface, let alone the dynamics of the force of mortality in the market-consistent measure, which is required to model longevity-linked options. However, we believe it is possible to extend the forward mortality rate framework, which would enable the pricing of mortality options, although we leave this for future work.

11.4 Conclusion

The valuation of longevity-linked liabilities and securities requires us to predict future rates of mortality. Modern solvency regulations and the gradual emergence of a market in longevity-linked securities require these predictions to incorporate market information, in order to give prices for different securities which are consistent with those observed in the marketplace. As many previous studies have shown, forward mortality models are ideally placed to achieve this.

We therefore believe that the answer to the titular question raised in Norberg (2010) - are forward mortality rates the way forward? - is yes. Nevertheless, it is important to take on board the criticisms of Norberg (2010) and to develop a framework specifically to model mortality rates, rather than borrow a pre-existing framework developed for interest rates and to define this framework using securities which do not depend on the idiosyncratic timing of individual deaths. This is because, with a properly developed framework, we can derive a model which is capable of capturing the complex dynamics of mortality rates, and so obtain consistency between models of the short and forward mortality rates.

In this study, we have developed such a framework for forward mortality rates which is based upon the dynamics of the force of mortality given by the class of age/period/cohort mortality models. This framework has the advantage of being easier to estimate from historical data than existing models, with market information being incorporated via a relatively parsimonious transformation of the forward mortality rates in the real-world measure. The framework is also very flexible, as it can be used in conjunction with many of the most popular models of the force of mortality, such as those proposed in Lee and Carter (1992) and Cairns et al. (2006a).

We have shown how market information can be incorporated into the model and used the resulting forward mortality surface to value a range of existing and proposed longevity-linked securities. All of the prices calculated from the same model are consistent with each other, as they are derived from the same forward surface of mortality. This allows for a unified approach to the valuation of a wide range of liabilities and longevity-linked securities.

Finally, we note that the main virtue of forward mortality models is their ability to specify the dynamics of the forward mortality surface and, hence, their applicability to the assessment and management of longevity risk. We develop these themes in the second part of this study, in Chapter 12. Together, these two studies show that the framework proposed can provide an integrated solution to many of the valuation and risk management problems in respect of longevity risk that are faced by life insurance companies.

11.A Identifiability and mortality forward rates

In Chapters 3 and 4, we discuss the identifiability issues in AP and APC mortality models, respectively. In particular, we find that almost all APC mortality models possess "invariant" transformations, i.e., transformations of the parameters of the model which leave the fitted mortality rates unchanged. In order to find a unique set of parameters, we impose a set of identifiability constraints on them. Typically, these are chosen to give a particular demographic significance to each term in the model. However, since any interpretation of demographic significance is subjective, it is important that our choice of identifiability constraints does not have any impact on any conclusions we draw about historical or projected mortality rates. For instance, we discuss in Chapters 3 and 4 how to ensure that projected force of mortality is independent of the choice of identifiability constraint.

It is also important that the forward mortality rate framework described in this study is independent of the choice of identifiability constraints used when fitting the underlying APC model to historical data. However, due to our definitions of the forward mortality rates in Equation 11.11, we see that $\nu^{\mathbb{P}}_{x,t}(\tau)$ in the real-world measure is automatically independent of the identifiability constraints if the distribution of $\mu_{x,\tau}$ is also independent of the identifiability constraints. We therefore do not need to do any additional work to ensure identifiability in the forward rates once the methods used to project the force of mortality are well-identified.

We also need to ensure that the forward mortality surface in the market-consistent measure is also independent of the choice of arbitrary identifiability constraints. This is mostly straightforward, as we see that Equation 11.31 depends upon the forward mortality rates in the real-world measure (which should be independent of the identifiability constraints for the reasons discussed above), the variances of the period and cohort functions (which are independent of the allocation of any levels and linear trends if the projection methods are well-identified, as discussed in Chapter 4) and the market prices of longevity risk. However, we note that if the model transformed using

$$\{\hat{\boldsymbol{\beta}}_x, \hat{\boldsymbol{\kappa}}_t\} = \{\left(A^{-1}\right)^\top \boldsymbol{\beta}_x, A\boldsymbol{\kappa}_t\}$$

then the market prices of risk are also transformed in the model to $\hat{\boldsymbol{\lambda}} = \left(A^{-1}\right)^\top \boldsymbol{\lambda}$. Hence we see that, not only are the values of the market prices of risk dependent upon the underlying APC model used for the force of mortality, they will also depend upon

the normalisation scheme and specification of the age function in the model, and so are not the same across all models which give the same fitted mortality rates.

11.B Impact of Jensen's inequality

In Section 11.2.2, it was argued that

$$\begin{aligned} {}_tP_{x,\tau} &= \mathbb{E}_\tau\left[\exp\left(-\sum_{u=1}^{t}\mu_{x+u,\tau+u}\right)\right] \\ &\approx \exp\left(-\sum_{u=1}^{t}\mathbb{E}_\tau\mu_{x+u,\tau+u}\right) \end{aligned} \tag{11.35}$$

due to the relatively low degree of variability in $\mu_{x,t}$, and hence it was shown in Section 11.2.2 that

$$\nu_{x,t}(\tau) \approx \mathbb{E}_\tau\mu_{x,t}$$

This assumption can be tested numerically, as follows.

For simplicity, we consider $P_{x,t} = \mathbb{E}_\tau \exp(-\mu_{x,t})$. Therefore

$$P_{x,t} = \mathbb{E}_\tau \exp\left(-\exp\left(\eta_{x,t}\right)\right)$$

In Section 11.2.3, we assume that

$$\eta_{x,t} \sim N(\mathcal{M}_{x,t}, \mathcal{V}_{x,t})$$

and therefore

$$\mathbb{E}_\tau\ \exp(-\mu_{x,t}) \approx \exp\left(-\mathbb{E}_\tau\mu_{x,t}\right) = \exp\left(-\exp\left(\mathcal{M}_{x,t} + 0.5\mathcal{V}_{x,t}\right)\right) \tag{11.36}$$

Holland and Ahsanullah (1989) discussed the log-log distribution, where X is such that

$$\ln(-\ln(X)) \sim N(\mathcal{M}, \mathcal{V})$$

We therefore see that $P_{x,\tau}(\tau)$ is given by the mean of the log-log distribution if $\eta_{x,t}$ is normally distributed. However, the moments of this distribution do not have a closed form solution. Holland and Ahsanullah (1989) showed that the r^{th} raw moment of the

distribution is given by

$$\mathbb{E}X^r = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-0.5x^2 - r \ \exp[\mathcal{M} + x\sqrt{\mathcal{V}}]\right) dx$$

which can be computed numerically.

From Section 11.2.3, we see

$$\begin{aligned}\mathcal{M}_{x,t} &= \alpha_x + \boldsymbol{\beta}_x^\top \mathbb{E}_\tau \boldsymbol{\kappa}_t + \mathbb{E}_\tau \gamma_{t-x} \\ \mathcal{V}_{x,t} &= \boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{V}ar_\tau(\gamma_{t-x})\end{aligned}$$

Hence we can use the results of Holland and Ahsanullah (1989) to compute $P_{x,t}$ numerically, without recourse to the approximation in Equation 11.36. Using this, we calculate

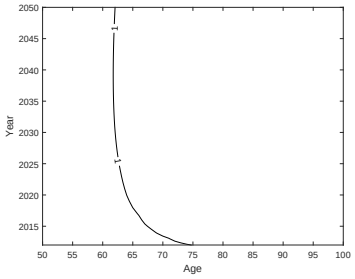
$$\begin{aligned}P_{x,t} &= \mathbb{E}_\tau \exp(-\mu_{x,t}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-0.5z^2 - \exp[\mathcal{M}_{x,t} + x\sqrt{\mathcal{V}_{x,t}}]\right) dz \qquad (11.37)\end{aligned}$$

numerically and compare it with the values assumed in Equation 11.36. This gives us a check on the accuracy of the approximation in Equation 11.36, which underpins the forward mortality framework.

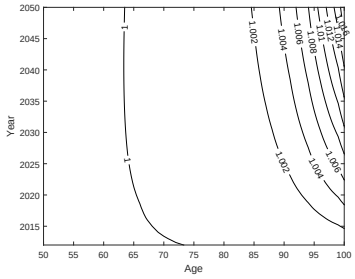
Figure 11.8 shows the ratio of the numerical value of $P_{x,t}$ calculated using Equation 11.37 and the approximate value calculated using Equation 11.36 for the five mortality models considered in this paper (in the real-world measure). We can that in the vast majority of cases, the difference that the assumption makes is less than 0.2% (i.e., ratios less than 1.002) and for no ages and years does the approximation make more than a 1.5% difference to the forward mortality rates. This is consistent with the projected mortality rates found in Figure 11.1, which also showed that forward mortality rates (using the approximation) were very close to those calculated using Monte Carlo simulations.

The mortality rates which are most affected by the approximation are those at the highest ages and the years of projection furthest into the future, which makes sense as these are the mortality rates with the greatest levels of uncertainty attached to them. However, they are also the least economically important, since any cashflows that would be affected by these mortality rates would be in respect of individuals who are very old (and so there is very little survivorship to these ages) and far into the future (which means that the present value of the affected cashflows would be very small due to discounting). This gives

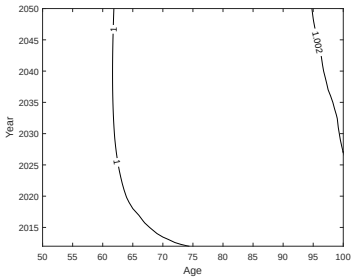
FIGURE 11.8: Impact of Jensen's inequality



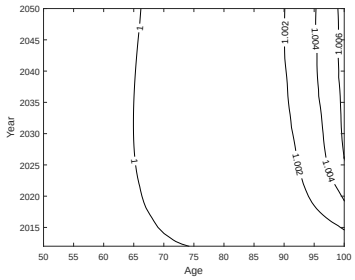
(A) Lee-Carter model



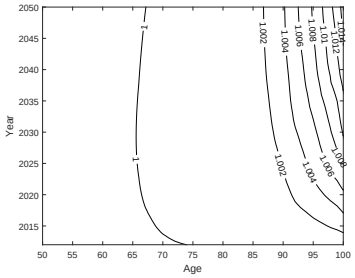
(B) CBDX model



(C) Classic APC model



(D) Reduced Plat model



(E) General procedure model

us reassurance that the approximation in Equation 11.35 does not systematically distort the results found using the forward mortality framework derived in this study, compared with those which could be found using an exact but considerably more complicated framework which does not make this assumption.

Chapter 12

Forward Mortality Rates in Discrete Time II: Longevity Risk Measurement and Management

12.1 Introduction

The first decade of the 21^{st} century has witnessed the realisation of the importance of longevity risk in the provision of retirement benefits and the emergence of new securities and derivatives, such as the longevity swaps and pension buy-ins discussed in Blake et al. (2013), to manage this risk. It has also witnessed a financial crisis and resulting recession, caused, in part, by new forms of financial securities and the faulty measurement and management of risk surrounding them. It is, therefore, of paramount importance that we do not make the same mistakes with the growing market for longevity risk that were made in the market for mortgage-backed securities. Consequently, it is vital to be able to measure and manage the risk in longevity-linked liabilities securities reliably and consistently.

Longevity risk is often defined as the risk that life expectancy increases at a faster rate than anticipated, or conversely, that mortality rates decrease faster than expected. However, in the context of liability-linked liabilities and securities, the major financial impact of longevity risk is not the difference between anticipated and actual mortality rates. Instead, it is the impact of changes in the expectations of future mortality rates that has the greatest impact on the valuation of these liabilities and securities.

Since the expectations of future mortality rates are forward-looking by definition, what is required for the measurement of longevity risk is a forward model for mortality rates. In Chapter 11, we developed such a forward mortality framework, based on the dynamics of the force of mortality given by age/period/cohort (APC) models in discrete time. We then demonstrated how such a model can be calibrated to market information, in order to price a range of longevity-linked liabilities and securities consistently, both with the market information we possess and with each other. Because market-consistent values are required for the liabilities under the Solvency II regulatory regime, as described in EIOPA (2014), such a forward mortality framework could also form the basis of an insurer's internal model for longevity risk.

In this chapter, we go beyond defining the surface of forward mortality rates at a single point in time to consider how this surface will change in future. These changes are driven by the dynamics for the parameters of the underlying APC mortality model and so are consistent with how the forward mortality surface was defined initially. Changes in the forward mortality rates then feed through into changes in the values of longevity-linked liabilities and securities, and so form the basis of the measurement of longevity risk. This is especially important in the context of modern regulatory regimes, such as Solvency II, where an accurate determination of the capital required to support different life insurance liabilities is a critical business issue. Since the forward mortality framework gives consistent values for both longevity-linked liabilities and securities, we can also use it to measure the impact of hedging strategies which attempt to manage longevity risk.

The structure of this chapter is as follows. We first consider how the forward surface of mortality will evolve over a one year period in Section 12.2 by examining the processes assumed to be generating the observed period and cohort parameters. This is then applied in Section 12.3 to examine the riskiness of annuity values using different risk measures and the impact of hedging liability values using simple longevity-linked securities. In Section 12.4, this analysis is extended to the measurement of longevity risk over multiple years, with a particular application to calculating the "risk margin" under the proposed Solvency II regulations and the numerical issues caused by this calculation. Finally, Section 12.5 concludes.

12.2 One-year updates of the forward mortality surface

The mortality forward rate framework discussed in Chapter 11 enables us to value longevity-linked liabilities and securities values in a market-consistent fashion. However, for many risk measurement purposes we are also interested in how these values change with time. There will be three components to such changes:

1. Changes in value due to changing conditions in financial markets not linked to longevity, for instance, due to changes in interest or inflation rate expectations. Changes in these quantities have been widely studied and a range of models have been developed for interest rates and inflation that could be used to deal with the impact of these changes on longevity-linked liabilities and securities values. Accordingly, we do not study the impact of these changes in this chapter.[1]

2. Changes due to new mortality data. Mortality data is released relatively infrequently, typically annually, and would be used to refit the underlying APC mortality model. Such changes will be considered further in this study.

3. Changes due to changing market longevity-risk preferences. These would result in changes in the values of traded securities not explainable in terms of new mortality data or changes in other non-demographic market indicators, and would be incorporated into the forward mortality rate model as time-dependent market prices of longevity risk, $\lambda^{(j)}(\tau)$. With the traded market in longevity-linked securities in a very early stage of development, there is no reliable information available to determine how these changes should be modelled. As Blake et al. (2006) said *"sophisticated assumptions about the dynamics of the market price of longevity risk are pointless"*, given the absence of market data to calibrate them. We therefore assume that the market prices for longevity risk are constant and do not consider them further.

[1] We also implicitly assume that processes governing the evolution of mortality rates are independent of other financial risks. This is in common with the majority of studies, such as Cairns et al. (2006b) and Bauer et al. (2008) and with the available evidence to date, as discussed in Loeys et al. (2007). Although there may be some situations where longevity risk is not independent of other financial risks in the real-world measure, as in the examples of Miltersen and Persson (2005), we believe that these situations are relatively extreme and are better considered by scenario analysis rather than through a stochastic model. Furthermore, Dhaene et al. (2013) show that independence between longevity risk and financial risks in the real-world measure does not automatically ensure independence in the market-consistent measure. However, more complicated models are required in order to allow for any dependence between longevity and investment risks, which require more market information for calibration. Therefore, we believe that the assumption of independence between longevity risk and other financial risks is necessary and justifiable at this early stage of development of the longevity risk market.

To investigate the second component of these changes, we are, therefore, interested in the random variables

$$\nu^{\mathbb{Q}}_{x,t}(\tau+1)|\mathcal{F}_\tau$$

i.e., the distribution of the forward mortality rates at $\tau+1$ conditional on information at time τ. This is equivalent to studying the "updating factors"

$$\frac{\nu^{\mathbb{Q}}_{x,t}(\tau+1)}{\nu^{\mathbb{Q}}_{x,t}(\tau)}$$

which underpins the models of Cairns (2007) and Zhu and Bauer (2011b).

In reality, the process of determining the forward surface of mortality would involve acquiring death counts and exposures to risk across all ages for year $\tau+1$, re-estimating the chosen mortality model with a revised dataset which included this new information to obtain new estimates of the various age, period and cohort parameters and then using these revised estimates within the framework of Chapter 11. However, this process is not practical for risk management purposes, as the process of generating new death counts and exposures to risk and refitting the model can be sufficiently time consuming that it is not viable to perform it thousands of times. Instead, we note the key new information which the additional data gives us:[2]

1. We can use the new data to estimate for the first time the value of $\boldsymbol{\kappa}_{\tau+1}$.
2. We can use the new data to re-estimate the cohort parameters, and so revise the old fitted cohort parameters, $\overline{\gamma}_y(\tau)$, to a new set of fitted cohort parameters, $\overline{\gamma}_y(\tau+1)$.

Accordingly, to avoid the need to simulate death counts and exposures for $\tau+1$ and refit the model, we instead generate new "observations" of $\boldsymbol{\kappa}_{\tau+1}$ and $\overline{\gamma}_y(\tau+1)$ based on the assumed time series dynamics which underlie the forward mortality framework. The procedures for doing this are discussed in Sections 12.2.1 and 12.2.2 for the period and the cohort functions, respectively.

In following this procedure, it is important to ensure that our updated forward mortality surface is "self-consistent", as defined in Zhu and Bauer (2011b), namely that *"that expected values of future forecasts should align with the current forecasts"*. This means that forward mortality rates should be martingales. Such a condition is similar to "no

[2] A similar line of reasoning can be found in Tan et al. (2014), which used the "time invariant" property of the period functions in some mortality models to investigate the hedging of longevity risk.

arbitrage" conditions in forward interest rate models. However, because the markets for longevity risk are not complete and are likely to involve a more diverse range of potential underlying securities,[3] we cannot rule out the possibility of arbitrage opportunities even in a self-consistent framework. Given the definition of the forward mortality rates in Equation 11.11, we note that[4]

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}{}_{\tau}\,\nu^{\mathbb{P}}_{x,t}(\tau+1) &= \mathbb{E}^{\mathbb{P}}{}_{\tau}\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\,\mu_{x,t} \\ &= \mathbb{E}^{\mathbb{P}}{}_{\tau}\,\mu_{x,t} \\ &= \nu^{\mathbb{P}}_{x,t}(\tau) \end{aligned} \tag{12.1}$$

by the tower property of conditional expectations. This means that real-world measure forward mortality rates are self-consistent in the real-world measure. We can verify this by considering the period and cohort functions separately, which is done in Section 12.2.1 for the period parameters and Appendix 12.A.1 for the cohort parameters.

A similar line of reasoning leads to

$$\mathbb{E}^{\mathbb{Q}}{}_{\tau}\,\nu^{\mathbb{Q}}_{x,t}(\tau+1) = \nu^{\mathbb{Q}}_{x,t}(\tau)$$

i.e., market-consistent forward mortality rates are self-consistent in the market-consistent measure. This result is verified algebraically in Appendix 12.A.2 and provides a useful and important check on the validity of the modelling approach and ensures that there are no internal contradictions.

For most of the practical risk management purposes in Section 12.3, what is of interest is how values of liabilities and securities change in the real-world measure (e.g., to find the one-in-200 real-world scenario under Solvency II). Since these values are calculated using market-consistent forward mortality rates, the value of liabilities and securities are not self-consistent in the real-world measure. However, this is not surprising and is similar to other results in finance.[5] Nevertheless, it will have a number of consequences for the behaviour of longevity-linked liabilities and securities, as discussed in the following sections.

[3] Such as longevity zeros (based on survivorship), q-forwards (based on probabilities of death), e-forwards (based on period life expectancy) and other securities based on bespoke indices.

[4] We adopt the convention that the subscript on operators $\mathbb{E}_{\tau}(.)$, $\mathbb{V}ar_{\tau}(.)$ or $\mathbb{C}ov_{\tau}(.)$ denotes conditioning on the information available at time τ, i.e., $\mathcal{F}_{\tau}$.

[5] For example, the Black-Scholes stock option price is a martingale in the risk-neutral measure by construction. When performing risk management on stock options in the real-world measure, the options prices will not be martingales (in general, we would expect to see the value of a call option increase with time, since the share price is expected to grow faster than the risk-free rate).

In Chapter 11, we used the forward mortality framework with a number of different APC models, including the Lee-Carter model (Lee and Carter (1992)), the classic APC model of Hobcraft et al. (1982) and the model developed in Chapter 9 using the "general procedure" (GP) of Chapter 5. In this chapter, we only use the GP model as it provides a good fit to the historical data and possesses most of the features of more complicated mortality models such as multiple age/period terms and a cohort term. However, it is important to note that the techniques we propose could be used in combination with any mortality model within the class of APC models discussed in Chapter 2.

12.2.1 Period parameters

Consider first the period functions. From Equation 11.16 and 11.17, we have

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\boldsymbol{\kappa}_t &= \boldsymbol{\kappa}_{\tau+1} + \mu \sum_{s=\tau+2}^{t} X_s \\
\mathbb{V}ar^{\mathbb{P}}_{\tau+1}\left(\boldsymbol{\kappa}_t\right) &= (t-\tau-1)\Sigma
\end{aligned}$$

Therefore, by generating a value of $\boldsymbol{\kappa}_{\tau+1}$ using the random walk with drift process underlying the projections, we can update the means and variances of the future period functions (and hence the forward surface of mortality) from those found at τ to a (stochastic) update at $\tau+1$:

$$\begin{aligned}
\boldsymbol{\kappa}_{\tau+1} &= \boldsymbol{\kappa}_{\tau} + \mu X_{\tau+1} + \boldsymbol{\epsilon}_{\tau+1} \\
\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\boldsymbol{\kappa}_t &= \boldsymbol{\kappa}_{\tau+1} + \mu \sum_{s=\tau+2}^{t} X_s \\
&= \boldsymbol{\kappa}_{\tau} + \mu X_{\tau+1} + \boldsymbol{\epsilon}_{\tau+1} + \mu \sum_{s=\tau+2}^{t} X_s \\
&= \boldsymbol{\kappa}_{\tau} + \mu \sum_{s=\tau+1}^{t} X_s + \boldsymbol{\epsilon}_{\tau+1} \\
&= \mathbb{E}^{\mathbb{P}}{}_{\tau}\boldsymbol{\kappa}_t + \boldsymbol{\epsilon}_{\tau+1} \\
\mathbb{V}ar^{\mathbb{P}}_{\tau+1}\left(\boldsymbol{\kappa}_t\right) &= (t-\tau)\Sigma - \Sigma \\
&= \mathbb{V}ar^{\mathbb{P}}_{\tau}\left(\boldsymbol{\kappa}_t\right) - \Sigma
\end{aligned}$$

Hence we see that the expectation of future period parameters changes by the innovation $\boldsymbol{\epsilon}_{\tau+1}$ for all future times, whilst the variance of the future period parameters reduces to

reflect that, at $\tau+1$, they will be projected for one fewer year than at τ.

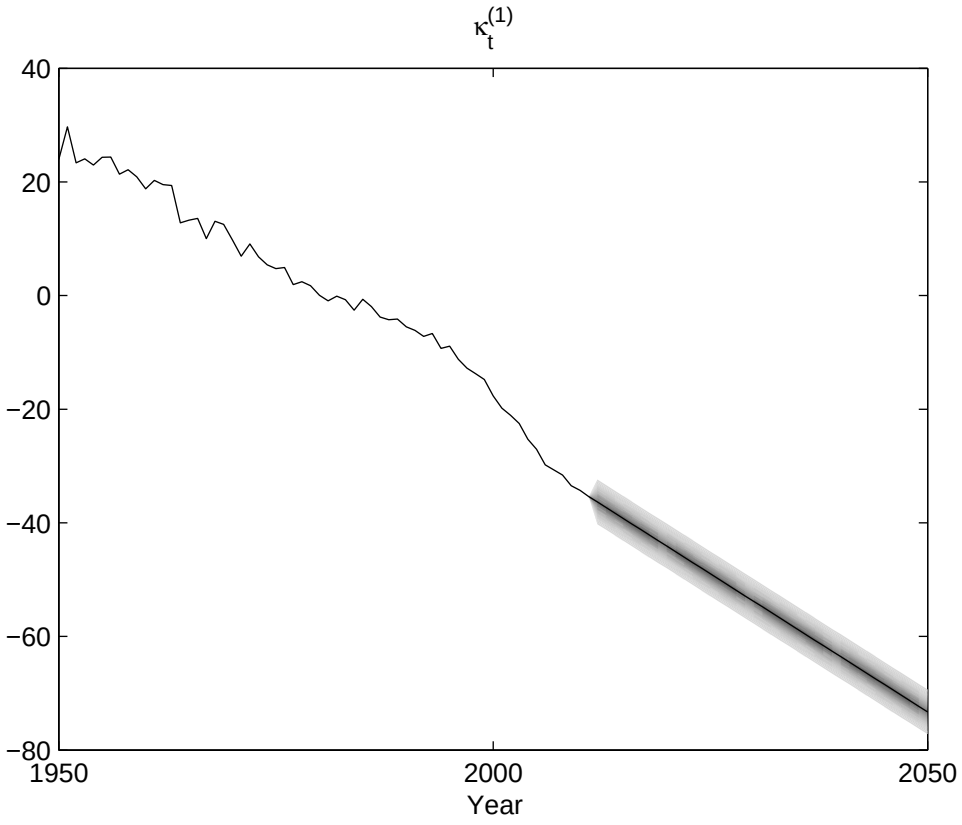


FIGURE 12.1: 95% prediction interval for $\mathbb{E}_{\tau+1}\kappa_t^{(1)}|\mathcal{F}_\tau$

Figure 12.1 shows the 95% prediction interval for $\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\,\kappa_t^{(1)}|\mathcal{F}_\tau$ from the GP model. As can be seen, it is the value of $\kappa_{\tau+1}^{(1)}$ which generates the uncertainty in the later period functions, which shift in parallel as a result of this new information.[6]

To demonstrate the impact of this update of the period functions on the forward mortality rates, we see that

$$\begin{aligned}
\nu_{x,t}^{\mathbb{P}}(\tau+1)|\mathcal{F}_\tau &= \exp\left(\alpha_x + \boldsymbol{\beta}_x^\top \mathbb{E}_{\tau+1}\boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x \mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x^\top\right)|\mathcal{F}_\tau \\
&= \exp\left(\alpha_x + \boldsymbol{\beta}_x^\top\left(\mathbb{E}_\tau \boldsymbol{\kappa}_t^\top + \boldsymbol{\epsilon}_{\tau+1}\right) + \frac{1}{2}\boldsymbol{\beta}_x^\top\left(\mathbb{V}ar_\tau\left(\boldsymbol{\kappa}_t\right) - \Sigma\right)\boldsymbol{\beta}_x\right)|\mathcal{F}_\tau \\
&= \exp\left(\boldsymbol{\beta}_x^\top \boldsymbol{\epsilon}_{\tau+1} - \frac{1}{2}\boldsymbol{\beta}_x^\top \Sigma \boldsymbol{\beta}_x\right)\nu_{x,t}^{\mathbb{P}}(\tau)
\end{aligned}$$

[6] Note that, as the drift of the random walk process, μ, is assumed to be known, the forward mortality framework does not allow for what was termed "recalibration" risk in Cairns et al. (2013), i.e., the risk that one year's new information will cause a reappraisal of the drift term. We leave the inclusion of recalibration risk in the framework as future work. This may understate the risk in long-term projections of mortality rates and forms a key difference between our results and those of Richards et al. (2014).

if the underlying mortality model of the mortality short rate does not possess a cohort term. Hence, generating random values of $\boldsymbol{\epsilon}_{\tau+1}$ (the time-series innovations for the period parameters) can therefore be used to update stochastically the forward mortality surface at $\tau+1$, conditional on information to time τ in a relatively straightforward fashion. In addition, we see that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}{}_{\tau}\,\nu^{\mathbb{P}}_{x,t}(\tau+1) &= \exp\left(\boldsymbol{\beta}^{\top}_{x}\mathbb{E}^{\mathbb{P}}\tau\boldsymbol{\epsilon}_{\tau+1} + \frac{1}{2}\boldsymbol{\beta}^{\top}_{x}\mathbb{V}ar_{\tau}(\boldsymbol{\epsilon}_{\tau+1})\boldsymbol{\beta}_{x} - \frac{1}{2}\boldsymbol{\beta}^{\top}_{x}\Sigma\boldsymbol{\beta}_{x}\right)\nu^{\mathbb{P}}_{x,t}(\tau)\\ &= \nu^{\mathbb{P}}_{x,t}(\tau)\end{aligned}$$

and, hence, the real-world forward mortality rates are martingales in the $\mathbb{P}$-measure as expected.

12.2.2 Cohort parameters

As discussed above, the impact of new data for year $\tau+1$ has a fundamentally different impact on the cohort parameters compared with the period parameters in a mortality model. For the period parameters, new data would allow us to estimate a value for $\boldsymbol{\kappa}_{\tau+1}$. To approximate this, we use the time series dynamics of the period functions to project $\boldsymbol{\kappa}_{\tau+1}$ stochastically, and use this to update the forward surface of mortality.

In contrast, new death count and exposure to risk data allows us to:

1. update the cohort parameters estimated by the model to allow for one additional observation on each cohort which is alive at $\tau+1$;

$$\overline{\gamma}_{y}(\tau) \to \overline{\gamma}_{y}(\tau+1) \quad \text{for } \tau+1-X \leq y \leq Y$$

2. estimate for the first time the cohort parameter for year of birth $Y+1$, i.e., $\overline{\gamma}_{Y+1}(\tau+1)$, which we did not have sufficient information to do the year before.

Unlike for the period functions, the new data does not give us a complete observation of any new, single year of birth. It is this fundamental difference in the information that new data provides that means that we need to adopt a fundamentally different approach when updating the cohort parameters in the forward mortality framework.

To explain why this is important, we need to first consider the problems with using more classical approaches to projecting the cohort parameters. In Chapter 11, we found that

classical approaches, such as those using ARIMA models, are not suitable in a forward mortality framework. This was because there is a discontinuity in the variance of the parameters when we move from the estimated parameters based on historical data to the projected parameters. This discontinuity would give rise to pricing anomalies. In the context of updating the forward mortality surface, we also find that using these classical approaches will lead to irregularities, as we now show.

Classical time series processes assume that the cohort parameters for which we have observations at time τ (up to and including γ_Y, say) are known with certainty and will not be revised and updated to reflect the new information received at $\tau + 1$. Instead, new information at $\tau + 1$ is assumed to be sufficient to estimate γ_{Y+1}. Thus, the use of classical approaches would give results analogous to the updating of the period parameters above, i.e., that we only need to project γ_{Y+1} stochastically to reflect the impact of new data. The pattern of updated cohort parameters which would be observed using such as model is shown in Figure 12.2.

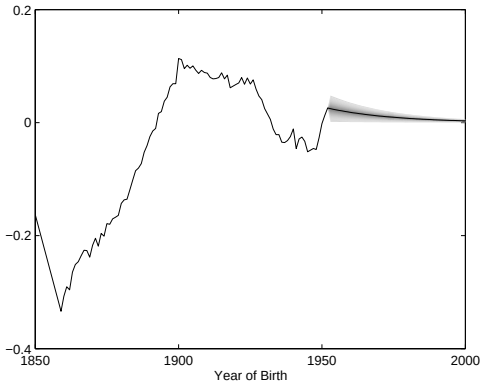


FIGURE 12.2: 95% prediction interval for the one-year update of projected γ_y using an AR(1) process

However, this is inconsistent with the impact new data would be expected to have, as discussed above. In addition, using these classical approaches generates unfeasible patterns of uncertainty in the forward mortality surface, with a sharp discontinuity between cohort parameters which are estimated from historical data and those which are projected, as discussed previously in Chapter 11.

In order to update the cohort parameters in a manner which is consistent with how they would actually update in response to new data, we instead need to use an approach which combines the time series dynamics of the cohort parameters with the partial observations we have of them to date. With such an approach, we can model the updating of this

partial information to reflect the impact of new data, and then combine this updated set of observations with the time series dynamics to revise our forecast cohort parameters. In Chapter 6, we developed a Bayesian modelling approach which can be used for this purpose. In particular, we assumed that we had two sources of information for estimating the "ultimate" cohort parameter, γ_y, which would only be known fully once all members of the cohort had died. These were the underlying time series dynamics for the cohort parameters, which acted as a prior assumption for their distribution, and the "interim" cohort parameters estimated by the mortality model, $\overline{\gamma}_y(\tau)$, which were based on partial information to time τ. Hence, the impact of new data on the cohort parameters can be modelled by generating updates of the estimated cohort parameters, $\overline{\gamma}_y(\tau+1)$, which reflect new observations of the relevant cohorts.

In Chapter 6, we assumed that the ultimate cohort parameters were generated by independent discrete packets, γ_y^x, for each age of observation for the cohort, i.e.,

$$\gamma_y = \sum_{x=1}^{X} d_x \gamma_y^x \tag{12.2}$$

where d_x is the proportion of the total cohort which dies at age x (assumed to be the same for all cohorts). However, at any specific time, we would only have received an incomplete set of observations of any cohort where members of that cohort were still alive, i.e., we would have received packets of information γ_y^x for $x \in [1, \tau - y]$ by time τ. These partial observations are combined to give us the estimated cohort parameters fitted by a mortality model based on data to time τ:

$$\underline{\gamma}_y(\tau) = \sum_{x=1}^{\tau-y} d_x \gamma_y^x \tag{12.3}$$

$$\overline{\gamma}_y(\tau) = \frac{1}{D_{\tau-y}} \underline{\gamma}_y(\tau) \tag{12.4}$$

where $D_x = \sum_{\xi=1}^{x} d_x$, i.e., the proportion of a cohort expected to die before age, x, as defined in Chapter 6.

Hence, the process of updating the cohort parameters to reflect new information for year $\tau+1$ is equivalent to generating new packets of information to represent the new observations of each of the still living cohorts at time $\tau+1$, and incorporating these into

the existing estimates of the cohort parameters at time τ

$$\underline{\gamma}_y(\tau+1) = \underline{\gamma}_y(\tau) + d_{\tau+1-y}\gamma_y^{\tau+1-y} \tag{12.5}$$

$$\begin{aligned}\overline{\gamma}_y(\tau+1) &= \frac{1}{D_{\tau+1-y}}\underline{\gamma}_y(\tau+1) \\ &= \frac{1}{D_{\tau+1-y}}\left[\underline{\gamma}_y(\tau) + d_{\tau+1-y}\gamma_y^{\tau+1-y}\right] \\ &= \frac{1}{D_{\tau+1-y}}\left[D_{\tau-y}\overline{\gamma}_y(\tau) + d_{\tau+1-y}\gamma_y^{\tau+1-y}\right]\end{aligned} \tag{12.6}$$

This can be compared to the results of a credibility analysis, as described in in Chapter 7 of Kaas et al. (2001), since the updated estimate of the cohort parameter is a weighted average of the previous estimate and the new observation of the cohort. Because of this, our ability to update the forward mortality surface for new cohort information rests on our ability to simulate new packets of information, $\gamma_y^{\tau+1-y}$. To do this, we know from Chapter 6 and the well-identified AR(1) process underlying the cohort parameters that

$$\gamma_y^x|\gamma_{y-1},\beta,\rho,\sigma^2 \sim N\left(\beta\tilde{X}_y + \rho(\gamma_{y-1} - \beta\tilde{X}_{y-1}), \frac{\sigma^2}{d_x}\right)$$

where β, $\tilde{X}_y$, ρ and σ^2 are defined in Chapters 6 and 11. However, the ultimate cohort parameter for year of birth $y-1$, γ_{y-1}, will not, in general, be known at time τ (as individuals born in year $y-1$ will still be alive), but we do know the distribution of γ_{y-1} at τ from Equations 11.20 and 11.21. Therefore, in order to find the distribution of $\gamma_y^{\tau+1-y}|\mathcal{F}_\tau$, we use Bayes Theorem and the distribution of γ_{y-1} to give

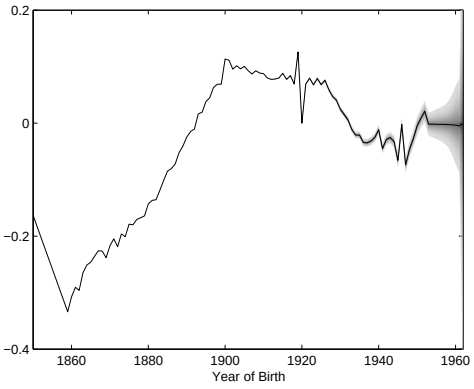
$$\gamma_y^{\tau+1-y}|\mathcal{F}_\tau,\beta,\rho,\sigma^2 \sim N\left(\beta X_y + \rho(M(y-1,\tau) - \beta X_{y-1}), \rho^2 V(y-1,\tau) + \frac{\sigma^2}{d_{\tau+1-y}}\right) \tag{12.7}$$

In addition, we assume

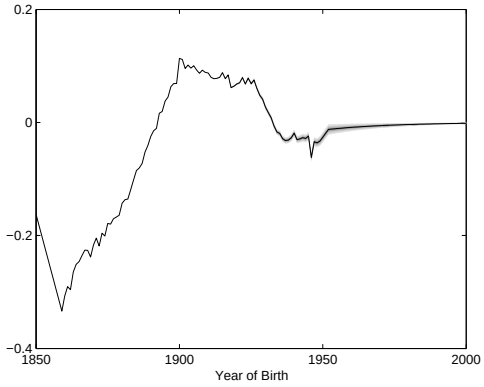
$$\mathbb{C}ov_\tau(\gamma_y^{\tau+1-y}, \gamma_{y-s}^{\tau+1-y+s}) = \rho^s\left[\prod_{r=0}^{s-1}(1 - D_{\tau+1-y+r})\right]\frac{\sigma^2}{d_{\tau+1-y+s}} \tag{12.8}$$

in order for the forward mortality rates to be self-consistent in the $\mathbb{P}$-measure, which is demonstrated in Appendix 12.A.1.

Hence, by generating new packets of information, γ_y^x, in respect of the cohorts that we would have observed in the new data for year $\tau+1$, we can update the values of $\overline{\gamma}_y(\tau)$ consistent with how they would update in response to actual new data.



(A) 95% prediction interval for the one-year update of interim cohort parameters, $\overline{\gamma}_y(\tau+1)$



(B) 95% prediction interval for the one-year update of the mean of the ultimate cohort parameters, $M(y,\tau+1)$

FIGURE 12.3: Updating the cohort parameters

To summarise, the process for updating the cohort parameters is:

1. generate new cohort information packets, $\gamma_y^{\tau+1-y}$ for $y \in [\tau+1-X, Y+1]$, randomly using the distribution in Equations 12.7 and 12.8;
2. update partial sums using Equation 12.6 without refitting the APC mortality model, to give $\overline{\gamma}_y(\tau) \to \overline{\gamma}_y(\tau+1)$;
3. use Equation 11.20 to find $M(y,\tau+1)$ (the updated estimate of the mean of the ultimate cohort parameters);
4. use Equation 11.21 to find $V(y,\tau+1)$ (the updated estimate of the variance of the ultimate cohort parameters);
5. use these to calculate $\nu^{\mathbb{P}}_{x,t}(\tau+1)$ in conjunction with the updated period parameters;
6. use Equation 11.31 to transform the real-world-measure forward mortality rates to the market-consistent measure, for use in valuing liabilities and securities.

The 95% prediction interval of the "interim" cohort parameters, $\overline{\gamma}_y(\tau+1)|\mathcal{F}_\tau$ is shown in Figure 12.3a, and the 95%prediction interval of the updated expectation of the ultimate cohort parameters, $M(y,\tau+1)|\mathcal{F}_\tau$ is shown in Figure 12.3b.[7] We observe the following:

- New data for $\tau+1$ does not update the cohort parameters for cohorts where we have assumed all members have died by time $\tau+1$, i.e., for $y \le \tau - X$.

[7]Note that, in $M(y,\tau)$ we use indicator variables to remove the large outliers due to the cohort anomalies in 1919/20 and 1946/47. This is because we believe them to be artefacts of the data collection process (see Richards (2008) and Cairns et al. (2014)), rather than genuine features of mortality for these cohorts.

- For years of birth $\tau + 1 - X \leq y \leq Y$, the new information would allow us to update the interim cohort parameter, $\overline{\gamma}_y(\tau)$, and hence the expectation of the ultimate cohort parameter, $M(y, \tau)$. The importance of this new information for the estimated cohort parameters is greater for more recent years of birth. This is reasonable, since the information received for year $\tau+1$ represents a greater share of the partial information received to this data for these years of birth. However, the Bayesian approach implies that the ultimate cohort parameters can be thought of as weighted averages of the prior distribution (given by the time series dynamics) and the partial information received by observing the cohorts to date, which is represented by $\overline{\gamma}_y$. For more recent years of birth, this approach gives greater weight to the prior distribution and less to the observations to date. Therefore, for recent years of birth, the impact of the new data updating the partial observations of the cohort (i.e., updating $\overline{\gamma}_y(\tau)$ to $\overline{\gamma}_y(\tau + 1)$) has only a limited impact on the distribution of the ultimate cohort parameters.

- We make our first estimate of the cohort parameters for year of birth $Y + 1$. This gives a very high variability for the estimated cohort parameter, $\overline{\gamma}_{Y+1}(\tau + 1)$, as this is based on very little information. However, since the Bayesian approach gives most weight to the time series dynamics for this cohort, this variability does not result in large changes in the expectation of the ultimate cohort parameter.

- For $y \geq Y + 2$, we still would not have sufficient observations to estimate $\overline{\gamma}_y(\tau+1)$. Hence the Bayesian approach gives no weight to the observations of the cohort (if any) to date and so the distribution of the ultimate cohort parameters for these cohorts is given entirely by the prior distribution, i.e., the time series dynamics. However, this prior distribution will have changed slightly because of the updated distributions of the ultimate cohort parameters for $y \leq Y + 1$. Since we have assumed that the cohort parameters follow an well-identified AR(1) process, updating the distribution of these parameters updates the prior distribution for the ultimate cohort parameters for $y \geq Y + 2$. However, these changes do not persist indefinitely and, instead, the impact of the new information decreases exponentially. This is reasonable, since we would not expect to update our estimates for the lifelong mortality features in respect of the cohort born in 2050 (say), based on observations of their parents and grandparents.

In these respects, the Bayesian framework has replicated what we would expect to see if we actually had new death counts and exposures for $\tau + 1$ and used them to refit the model. In addition, in Appendix 12.A.1, we check to ensure that the Bayesian framework for the cohort parameters gives self-consistent forward mortality rates in the real-world

measure.

Cohort effects are a feature of many of the more recent mortality models in use, and their robust estimation is of vital importance in the calculation of liabilities, such as annuities, and many of the longevity-linked securities which have been proposed. However, as discussed in Chapter 6, the projection of cohort parameters is difficult, and made more complicated by the nature of the partial information we have regarding them at any specific date. In part because of this, the forward mortality models proposed to date, such as those in the Heath-Jarrow-Morton framework in Barbarin (2008), Bauer et al. (2008) and Tappe and Weber (2013), the semi-parametric factor model of Zhu and Bauer (2011a,b, 2014), or the Olivier-Smith model developed in Olivier and Jeffrey (2004), Smith (2005), Cairns (2007) and Alai et al. (2013), have not been able to incorporate cohort effects. We believe that a key advantage of the forward mortality framework developed in Chapter 11 and in this study is that it can give biologically reasonable[8] dynamics for the forward surface of mortality, as it is based on the dynamics of APC models of the mortality hazard rate, which are well understood and easy to estimate from historical data. Since cohort parameters are an important feature of such models, we believe that the successful application of the forward mortality framework proposed in Chapter 11 and which will be used in the present study for risk management purposes is, ultimately, dependent upon using the Bayesian approach of Chapter 6.

12.3 One-year risk measurement and management

Based on the results of Section 12.2, we are able to generate random realisations of the forward mortality surface, which can then be used to value longevity-linked liabilities and securities. Doing so enables us to model how these values might change, which forms a key component in the measurement and management of longevity risk.

12.3.1 Annuity values

We begin by investigating the impact of the change in the forward mortality surface on the value of an annuity over a one-year period. Annuity values at each age, x, are

[8] Introduced in Cairns et al. (2006b) and defined as "*a method of reasoning used to establish a causal association (or relationship) between two factors that is consistent with existing medical knowledge*".

calculated as

$$a_x(\tau) = \sum_{t=0}^{\infty} {}_tP^{\mathbb{Q}}_{x,\tau}(\tau) B(\tau, \tau + t) \tag{12.9}$$

where ${}_tP^{\mathbb{Q}}_{x,\tau}(\tau)$ is the market-consistent forward survival probability from time τ to time $\tau + t$ (as evaluated at time τ), as defined in Chapter 11 and used in Equation 11.6, and $B(\tau, \tau + t)$ is the price at time τ of a risk-free zero coupon bond maturing at time $\tau + t$.[9] For these and all future calculations, we assume a constant risk free real rate of interest of 1% p.a. and extrapolate forward mortality rates beyond the maximum age in the data, $X = 100$, using the topping out procedure of Denuit and Goderniaux (2005).

This assumes that the lives on which the annuities are written are not systematically different from the national population, data for which was used to calibrate the forward mortality surface. Accordingly, we do not allow for potential basis risk in our annuity portfolio. We leave to future work the extension of the forward mortality framework to include basis risk, for example, using the relative modelling approaches of Villegas and Haberman (2014) or Chapter 9. However, the results of Chapter 9 indicate that the impact of basis risk on systematic longevity risk may be limited in many situations.

In order to assess the longevity risk in annuities over a one-year period, we first need to update the forward surface of mortality to time $\tau + 1$ using the techniques of Section 12.2 and then use this updated surface to calculate updated annuity values. These are given by

$$a_x(\tau + 1) = \sum_{t=0}^{\infty} {}_tP^{\mathbb{Q}}_{x,\tau+1}(\tau + 1) B(\tau + 1, \tau + 1 + t) \tag{12.10}$$

However, a direct comparison between these updated annuity values and those in Equation 12.9 is not valid. $a_x(\tau+1)$ is not directly comparable to $a_x(\tau)$, since it relates to the cohort born in $\tau + 1 - x$ as opposed to the cohort born in $\tau - x$. If, instead, one tries to compare $a_{x+1}(\tau + 1)$ with $a_x(\tau)$ (which do relate to the same cohort), we note that this comparison is also not valid, since the former includes one fewer year of benefits and is discounted to a different point in time compared with the latter. Consequently, we must

[9] We therefore see that an annuity is equal to a portfolio of longevity zeros, as defined in Blake et al. (2006) and used in Chapter 11.

be very careful in any comparisons that we make and compare $a_x(\tau)$ with[10]

$$B(\tau, \tau+1)_1 p_{x,\tau} \left(1 + a_{x+1}(\tau+1)\right) \tag{12.11}$$

Doing so values the same set of cashflows for the same cohort, discounted to the same point in time and therefore ensures that the two quantities are comparable. The difference between them arises from:

1. replacing the time τ market-consistent forward mortality rates in year $\tau+1$ with simulated "observed" rates for that year; and
2. replacing the time τ market-consistent forward mortality rates in years $t \geq \tau+2$ with the time $\tau+1$ market-consistent forward mortality rates for the same years.

Hence the only differences arise from changes arising from the changing forward surface of mortality and, therefore, they solely reflect longevity risk.

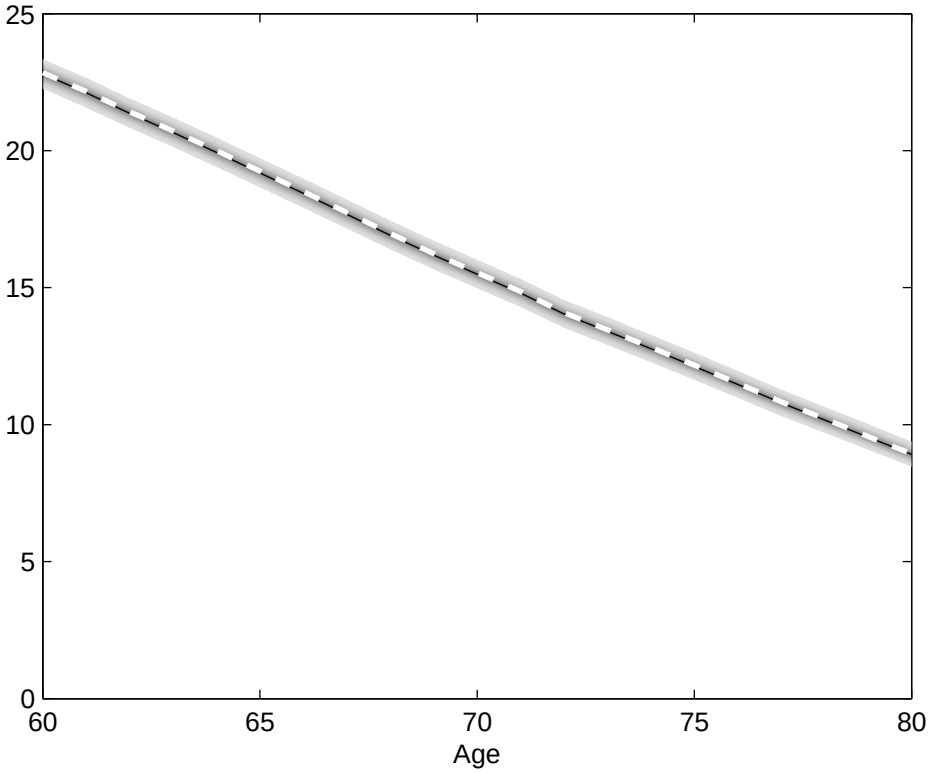


FIGURE 12.4: Projected annuity values at different ages at $\tau+1$

[10] In Equation 12.11 and subsequently, ${}_tp_{x,\tau}$ is the realised probability that an individual aged x at τ has survived to age $x+t$ at $\tau+t$

Figure 12.4 shows the 95% fan chart of simulated annuity values at different ages in one year's time. The coefficients of variation[11] of the projected annuity values increase with age, from around 1.4% of the current annuity value at age 60 to approximately 2.6% at age 80.

Figure 12.4 also shows the time τ annuity values, $a_x(\tau)$, as a dashed white line. It, therefore, illustrates that $\mathbb{E}^{\mathbb{P}}{}_{\tau}\ a_x(\tau+1) \approx a_x(\tau)$. However, it is important to note, however, that $\mathbb{E}^{\mathbb{P}}{}_{\tau}\ a_x(\tau+1) \neq\ a_x(\tau)$, i.e., the annuity values are not martingales in the real-world measure. The reason for this is that $a_x(\tau+1)$ is calculated using market-consistent forward mortality rates at time $\tau+1$, which are themselves not martingales in the real-world measure, as discussed in Section 12.2.

In Chapter 11, we said that the marginal participant in the market for longevity-linked securities would probably be a life insurer seeking to hedge longevity risk. Such a life insurer would be averse to longevity risk, and so, we expected that the market-consistent forward mortality rates would be lower than those in the real-world measure

$$\nu^{\mathbb{Q}}_{x,t}(\tau) \leq \nu^{\mathbb{P}}_{x,t}(\tau)$$

Thus, we expect to replace the expected survival probabilities for the period $[\tau, \tau+1)$ under the market-consistent measure with their projected values in the real-world measure, which are lower on average, i.e.,

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}{}_{\tau}{}_1p_{x,\tau} &= \mathbb{E}^{\mathbb{P}}{}_{\tau} \exp\left(-\mu_{x,\tau+1}\right) \\
&= \exp\left(-\nu^{\mathbb{P}}_{x,\tau+1}(\tau)\right) \\
&< \exp\left(-\nu^{\mathbb{Q}}_{x,\tau+1}(\tau)\right) = {}_1P^{\mathbb{Q}}_{x,\tau}(\tau)
\end{aligned}$$

Therefore, we find $\mathbb{E}^{\mathbb{P}}{}_{\tau}\ a_x(\tau+1) < a_x(\tau)$ across ages, indicating that annuity values would be expected to fall. In simulations, we find this has an impact of around 1% of the value of an annuity. In an insurance context, this would give an "expected return" due to the "release of reserves" in respect of the annuity, caused by having held reserves for the policy higher than the expected value of the benefits in the real-world measure. This expected return on longevity-linked liabilities and securities has important consequences, which will impact the measurement of risk in liabilities and longevity-linked securities,

[11] The standard deviation of the annuity value divided by its expectation.

as discussed in the following sections.

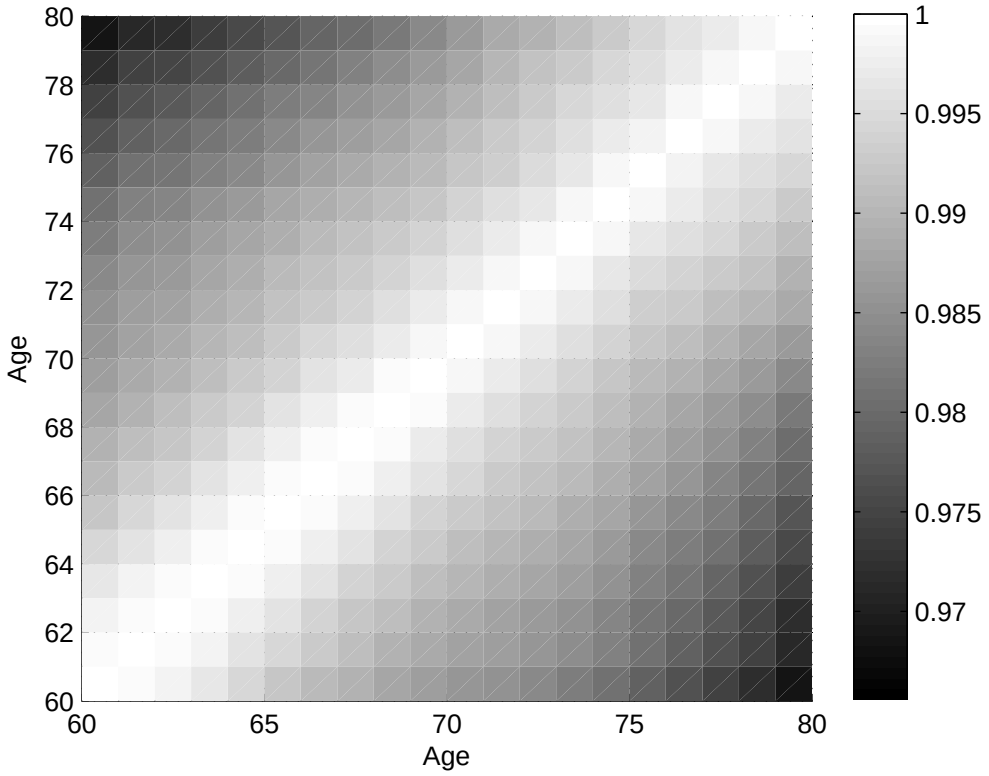


FIGURE 12.5: Correlations between annuity values at different ages at $\tau + 1$

In addition to looking at the annuity values at different ages in isolation, we also need to assess their dependence upon each other in order to achieve a full assessment of the longevity risk in our illustrative annuity book. To do this, Figure 12.5 shows the correlations between annuity values at different ages. From this, we see that there is substantial correlation between annuity values at different ages, typically between 95% and 100%. This is due to the structure of the underlying APC mortality model, since the evolution of the forward surface of mortality over the year is driven by the same few factors, namely the three age/period terms with a limited contribution from the cohort term. This leads, in turn, to relatively low diversification of longevity risk across different ages. In contrast, there could be apparently large benefits in risk reduction due to "natural hedging", i.e., writing life assurance policies as the value of these would be expected to be negatively correlated with annuity values under longevity risk, as discussed in Cox and Lin (2007). However, as argued in Zhu and Bauer (2014), these benefits are largely model dependent, although these criticisms can be partly assuaged by using APC mortality models with a sufficient number of terms to fully capture the dynamics of mortality.

However, for many risk measurement purposes, it is not sufficient to simply look at expectations, standard deviations and correlations of the annuity values. Instead, we need to use more sophisticated risk measures.

12.3.2 Risk measures

Numerous different risk measures are used in practice to quantify the riskiness of liabilities and portfolio values, many of which are discussed in Denuit et al. (2005) and Dowd et al. (2006b). Amongst these, some of the most commonly used risk measures are the "value at risk" (VaR) and the "tail value at risk" (TVaR). For a risk, X_1, occurring at time one, these are defined as

$$\text{VaR}(X_1;\alpha) = F_X^{-1}(1-\alpha) \tag{12.12}$$

$$\text{TVaR}(X_1;\alpha) = \mathbb{E}^{\mathbb{P}}\left[X_1 | X_1 \geq \text{VaR}(X_1;\alpha)\right] \tag{12.13}$$

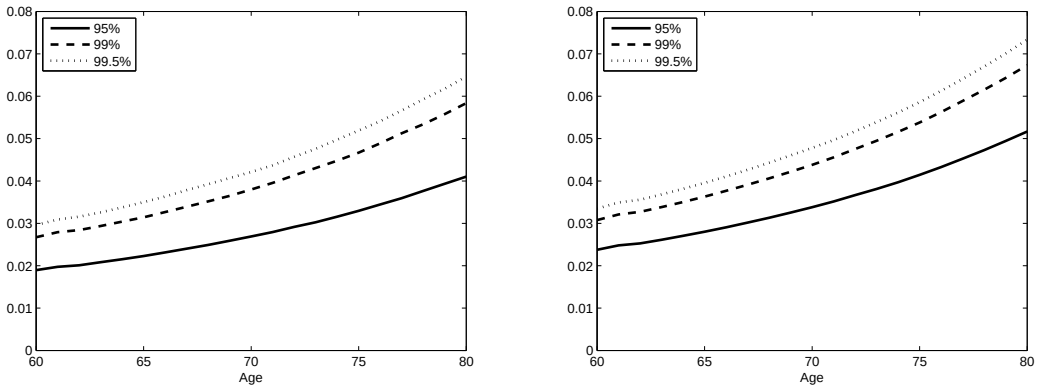
where α is the significance level of the risk measure and F_X is the cumulative distribution function for X_1 in the real-world measure, $\mathbb{P}$.[12] The value at risk can therefore be thought of as the loss observed $100\alpha\%$ of the time, whilst the tail value at risk can be interpreted as the expected value of the worst $100\alpha\%$ of the loss distribution (and hence it is also called the expected shortfall). Whilst the value at risk has numerous drawbacks as a risk measure, such as not being "coherent" as discussed in Denuit et al. (2005) and Dowd et al. (2006b), it remains widely used in practice as a benchmark for risk management, and is widely incorporated into regulations. The tail value at risk is coherent in the sense of Denuit et al. (2005), and also can be felt to give a more reasonable measure of the tail risk in a portfolio as it takes into consideration the distribution of the risk in the tail of the distribution, rather than merely the α^{th} quantile of this distribution.

For comparison purposes, rather than use VaR and TVaR directly, we define the "economic capital" as in Denuit et al. (2005) by

$$\text{EC}_{\varrho}(X_1;\alpha) = \varrho(X_1;\alpha) - \mathbb{E}^{\mathbb{P}}X_1 \tag{12.14}$$

where ϱ is the risk measure being used (i.e., VaR or TVaR). The economic capital therefore represents the capital required by an insurer to cover unexpected losses on risk X_1. This definition, using the value at risk forms the basis of the Solvency Capital Requirement (SCR) under the Solvency II regulatory capital regime, as discussed in EIOPA

[12] We assume that increasing X_1 corresponds to larger losses. In addition, as we will only deal with continuous X_1, the tail value at risk is equivalent to the conditional tail expectation. See Chapter 2 of Denuit et al. (2005) for more discussion of risk measures.



(A) VaR (B) $TVaR$

FIGURE 12.6: Economic capital ratios for annuity values at different ages

(2014) and below.

In this context, for comparison purposes, it is useful to go beyond this definition and compare "economic capital ratios" (ECRs) defined as follows

$$\begin{aligned} \mathrm{ECR}_{\varrho}(X_1;\alpha) &= \frac{\mathrm{EC}_{\varrho}(X_1;\alpha)}{X_0} \\ &= \frac{\varrho(X_1;\alpha) - \mathbb{E}^{\mathbb{P}}{}_{\tau}X_1}{X_0} \end{aligned} \tag{12.15}$$

where X_0 is the value of the risk at time zero. In the context of an insurer, the ECR can generally be thought of as referring to the amount of economic capital required per unit of "best-estimate" liability.

12.3.3 Risk measurement and management

12.3.3.1 Liabilities

Previously, we used the forward mortality framework to find the distribution of annuity values (controlling for the impact of benefits being paid, etc) updated to reflect an additional year of information. Consequently, we can use this distribution with the risk measures discussed in Section 12.3.2 to give a more detailed measurement of the longevity risk in annuity policies.

Figure 12.6a shows the $\mathrm{ECR}_{\mathrm{VaR}}$ for annuity values at different ages at the 95%, 99% and 99.5% levels (i.e., corresponding to one-in-20, one-in-100 and one-in-200 year events),

whilst Figure 12.6b shows ECR_{TVaR} for the annuities at the same levels. It is interesting to see that more capital (as a percentage of the best estimate liability) is required in respect of longevity risk for annuities for older individuals. This is despite the fact that annuities for these individuals are of shorter duration and therefore less subject to longevity risk. However, this is offset by the fact that annuities for older individuals have lower expected value, so the total economic capital will be lower than for younger-age annuities.

We believe this is because the primary impact of new data in our forward mortality model is to update the mortality rates observed in year $\tau+1$, which gives a broad impact across most ages. It is therefore interesting to compare these results with those presented in Richards et al. (2014), which showed smaller economic capital ratios for annuities at higher ages from a model that focuses primarily on extreme changes in the trend rate of improvement in mortality rates. This longevity trend risk was also called recalibration risk in Cairns et al. (2013), and we leave its inclusion in our forward mortality framework to future work.

The 99.5% VaR for longevity-linked liabilities is of particular interest to life insurance companies as it is used in the definition of the Solvency Capital Ratio (SCR) in the Solvency II regulatory requirements. These are set by the European Insurance and Occupational Pensions Authority (EIOPA) and are due to be implemented in 2016 for all insurance companies based in the EU (see also Stevens et al. (2010) and Bauer et al. (2012)). The liabilities side of the Solvency II balance sheet, described in EIOPA (2014), can be considered of consisting of two elements:

1. the "Technical Provisions", corresponding *"to the current amount undertakings would have to pay if they were to transfer their (re)insurance obligations immediately to another undertaking"* (EIOPA (2014, TP.1.1.)); and

2. the "Solvency Capital Ratio" (SCR) reflecting the additional capital required to protect against unanticipated risks, calculated as *"the Value-at-Risk of the basic own funds*[13] *of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period"* (EIOPA (2014, SCR.1.9.)).

In Chapter 11, we argued that the forward mortality framework could be used by a life insurer as an internal model to value its liabilities in a market-consistent fashion, and hence provide a valuation of the technical provisions described above. Together with the

[13] Defined as the difference between the assets and the liabilities in EIOPA (2014, SCR.1.6.).

results above, we therefore see that the forward mortality framework could be used to calculate both parts of the Solvency II balance sheet and, hence, act as an internal model for a life insurer with respect to its longevity risk.

To illustrate, we consider a stylised annuity book, consisting of annuities written on male lives equally distributed across ages 60 to 80. This liability profile has also been heavily simplified, as real annuity books are likely to include policyholders of both sexes[14] and different socio-economic backgrounds.[15] Nevertheless, it is sufficient to illustrate many of the advantages of the forward mortality framework and we will form the basis for our valuation of the components of the Solvency II balance sheet.

Technical provisions

Because there is no actively-traded market in longevity risk, it is impossible to accurately determine the technical provisions in a genuinely market-consistent fashion. There are two potential ways around this:

1. Construct a market-consistent measure that is somewhat subjective, perhaps via the inclusion of "internal" market information in the manner described in Chapter 11. The future benefit payments can then be valued in this measure to give a value for the technical provisions which is broadly market-consistent.

2. Use the real-world measure to value the future benefits payments, since this gives an objective value for them. However, EIOPA (2014) requires that, under this approach, the technical provisions would consist of this real-world value plus a "risk margin" to proxy for the additional cost of transferring the liabilities to a third-party. The calculation of the risk margin is complicated, and is discussed further in Section 12.4.2.

Using the market-consistent approach, the value of the future benefits (and hence the technical provisions) at time τ is calculated as

$$\mathcal{L}(\tau) = \sum_{x=60}^{80} a_x(\tau) \tag{12.16}$$

[14] Necessitating a multi-population model for the evolution of mortality, such as the one discussed in Chapter 8.

[15] As discussed in Villegas and Haberman (2014) which necessitates some form of individual risk scaling of the sort used in Chapter 10.

and the value at time $\tau + 1$ is

$$\mathcal{L}(\tau + 1) = \sum_{x=60}^{80} B(\tau, \tau + 1)_1 p_{x,\tau} \left(1 + a_{x+1}(\tau + 1)\right) \tag{12.17}$$

(both in notional currency units). Using this definition for the liabilities at $\tau + 1$ makes comparing $\mathcal{L}(\tau)$ and $\mathcal{L}(\tau + 1)$ more straightforward, in the same manner as was done above for the annuity values.

In contrast, using the real-world plus risk margin approach gives annuity values at age x and time τ of

$$\begin{aligned} a_x^{\mathbb{P}}(\tau) &= \sum_{t=0}^{\infty} {}_t P_{x,\tau}^{\mathbb{P}}(\tau) B(\tau, \tau + t) \\ &= \sum_{t=0}^{\infty} \exp\left(- \sum_{s=1}^{t} \nu_{x+s,\tau+s}^{\mathbb{P}}(\tau) \right) B(\tau, \tau + t) \end{aligned} \tag{12.18}$$

(compared to Equation 12.9), with similar modifications to Equations 12.16 and 12.17.

Using these approaches, we find values for the future liabilities of 331.4 for the market-consistent approach and 314.2 (both in notional currency units) using the real-world approach - a difference of 5.2%. This difference should be compensated for by the risk margin, as discussed in Section 12.4.2.

Solvency Capital Ratios

The difference in approach used for the valuation of the technical provisions also has consequences for the calculation of the SCR, since the "basic own funds" of the insurer depends upon the value of the technical provisions. However, the definition of the SCR is not precise, and there exist multiple potential interpretations of "basic own funds" which lead to subtly different values of the SCR for any given set of technical provisions - see Christiansen and Niemeyer (2014) for a more complete discussion of this issue. For instance, it is common to interpret basic own funds as the value of net assets (i.e., assets minus liabilities), as used in Stevens et al. (2010), and denoted as $\mathcal{N}(\tau)$ at time τ. Using this definition gives the following expression for the SCR

$$SCR(\tau) = \text{VaR}(\mathcal{N}(\tau) - B(\tau, \tau + 1)\mathcal{N}(\tau + 1) | \mathcal{F}_\tau; 99.5\%) \tag{12.19}$$

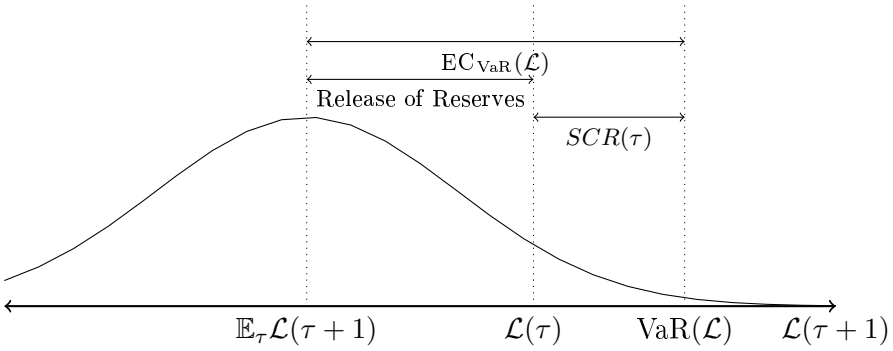


FIGURE 12.7: Decomposition of the SCR

In this study, we are interested only in the longevity risk in the liabilities, and thus assume that the assets are invested in riskless securities and so investment risk does not contribute to the SCR. Therefore, using the definition of the liabilities (calculated using either a market-consistent or real-world approach) in conjunction with Equation 12.21 gives

$$SCR(\tau) = \text{VaR}(\mathcal{L}(\tau+1) - \mathcal{L}(\tau)|\mathcal{F}_\tau; 99.5\%) \tag{12.20}$$

This can be decomposed as

$$\begin{aligned} SCR(\tau) &= \text{VaR}(\mathcal{L}(\tau+1)|\mathcal{F}_\tau; 99.5\%) - \mathcal{L}(\tau) \\ &= \Big(\text{VaR}(\mathcal{L}(\tau+1)|\mathcal{F}_\tau; 99.5\%) - \mathbb{E}^{\mathbb{P}}\left[\mathcal{L}(\tau+1)|\mathcal{F}_\tau\right]\Big) - \Big(\mathcal{L}(\tau) - \mathbb{E}^{\mathbb{P}}\left[\mathcal{L}(\tau+1)|\mathcal{F}_\tau\right]\Big) \\ &= \text{EC}_{\text{VaR}}(\mathcal{L}(\tau+1)|\mathcal{F}_\tau; 99.5\%) - \Big(\mathcal{L}(\tau) - \mathbb{E}^{\mathbb{P}}\left[\mathcal{L}(\tau+1)|\mathcal{F}_\tau\right]\Big) \end{aligned}$$

Consequently, we see that the common definition of the SCR consists of two parts:

1. the economic capital required to protect against unexpected longevity shocks at the one-in-200 level less
2. the expected release of reserves for the year, $\mathcal{L}(\tau) - \mathbb{E}^{\mathbb{P}}\left[\mathcal{L}(\tau+1)|\mathcal{F}_\tau\right]$.

This is illustrated in Figure 12.7.

We expect the release of reserves to be positive since the market-consistent measure is anticipated to project higher mortality rates than expected in the real-world measure, as discussed in Section 12.3.1 and, therefore, it will tend to offset the economic capital required to protect against risk. The magnitude of this release of reserves depends strongly on the specification of the market-consistent measure. Since the market-consistent measure, $\mathbb{Q}$, used in both Chapter 11 and this chapter is largely illustrative, due to the

absence of genuine market information on the prices of longevity-linked securities, we do not wish the details of its construction to bias our results. Consequently, we choose to define the SCR as

$$\text{SCR}(\tau) = \text{EC}_{VaR}(\mathcal{L}(\tau+1)|\mathcal{F}_\tau; 99.5\%) \tag{12.21}$$

i.e., the economic capital alone. Using the market-consistent approach to calculate the liability value, we find an SCR of 12.8, i.e., 3.9% of the value of the technical provisions.

In contrast, if a real-world approach is used to value the liabilities, we see that there is no release of reserves since

$$\mathbb{E}^{\mathbb{P}}{}_\tau\mathcal{L}^{\mathbb{P}}(\tau+1) \approx \mathcal{L}^{\mathbb{P}}(\tau)$$

i.e., the liability value in the real-world measure is almost a martingale, because the forward mortality rates are martingales in the real-world measure.[16]

Using the real-world approach to calculate the liability value, we find an SCR of 12.6, i.e., 4.0% of the best-estimate liability value. It is interesting to note that the nominal value of the SCR using the best-estimate liabilities is not significantly different from the value calculated using the market-consistent liabilities. This is because the change of measure does not introduce any additional uncertainty into the liabilities, and hence the nominal magnitude of their riskiness is the same.

In the context of using the forward mortality framework as an internal model under Solvency II, it is also interesting to compare the 99.5% economic capital ratios derived from the forward mortality framework with the "standard model" approach under Solvency II. This proposes that the SCR for longevity risk should be valued by assuming the probability of death at each age and for all time periods is reduced by 20% from what is expected under a best-estimate scenario. These stressed mortality rates are then used to value the liabilities EIOPA (2014, SCR.7.25). Figure 12.8 shows the SCRs found by such an approach and compares them with those found using the forward mortality framework (using the liabilities on a real-world basis).

[16] The difference between $\mathbb{E}^{\mathbb{P}}{}_\tau\mathcal{L}^{\mathbb{P}}(\tau+1)$ and $\mathcal{L}^{\mathbb{P}}(\tau)$ arises due to Jensen's inequality in a similar fashion to the approximation used in the definition of the forward mortality rates themselves in Chapter 11. However, the results of Chapter 11 show that this is likely to be negligible across all ages and years of interest.

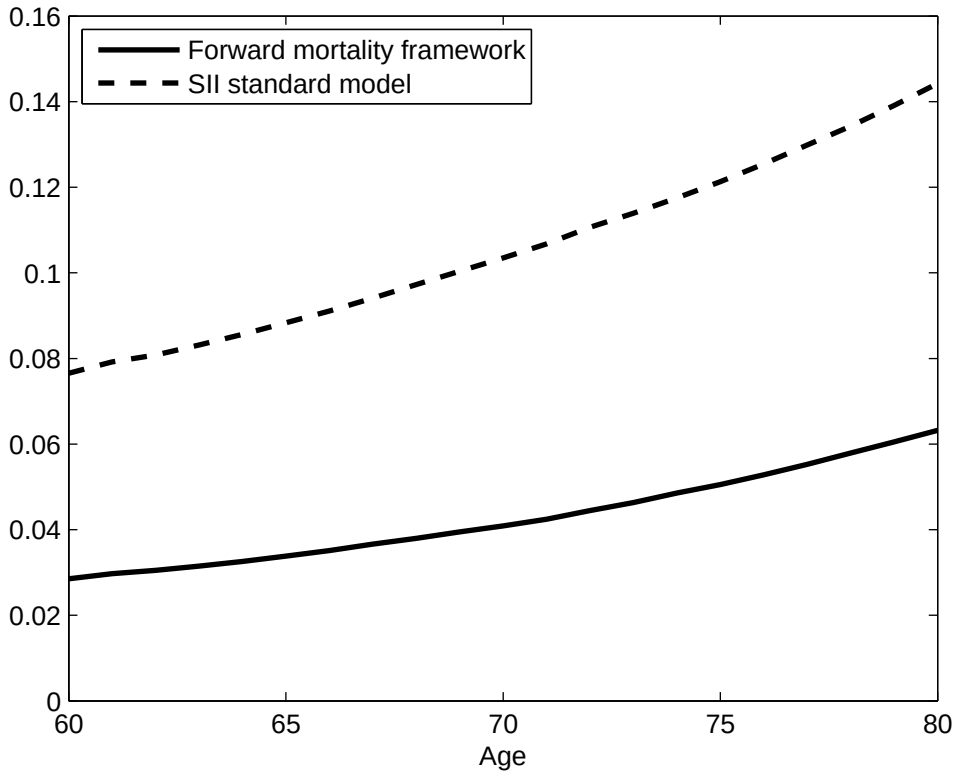


FIGURE 12.8: SCRs for annuities at different ages using the forward mortality framework and the Solvency II standard model

As can be seen, the Solvency II standard model for longevity risk overstates the required capital significantly compared with using the forward mortality framework - more than doubling the SCR for an annuity at most ages. This is comparable to the results found by other authors, such as Börger (2010), Nielsen (2010) and Richards et al. (2014) using a range of different models. In addition, the approach adopted by the Solvency II standard model, i.e., a one off reduction in mortality rates occurring immediately and remaining constant in time, is inconsistent with the nature of longevity risk, which is a long-term risk which increases over time.

These three approaches to calculating the technical provisions and SCR (using the market-consistent approach, the real-world approach and the standard model) are compared in Table 12.1. However, it is important to note that the relatively low value of the liability value found using the real-world and standard model approaches will be compensated for by the risk margin, which is considered in greater detail in Section 12.4.2.

Approach	Liability value	SCR(τ)	SCR as % of Liabilities
Market-consistent	331.4	12.8	3.9%
Real-world	314.2	12.6	4.0%
Standard model	314.2	31.6	10.0%

TABLE 12.1: Liability values and SCRs using difference approaches

12.3.3.2 Longevity-linked securities

In Chapter 11, the forward mortality framework was used to value a number of potential longevity-linked securities. For capital efficiency, most of these have taken the form of forward contracts, written on various indices of mortality. A number of different mortality indices for use in forward contracts have been proposed to date:

- q-forwards: as discussed in Coughlan et al. (2007b), these are forward contracts on future probabilities of death, $q_{x,t}$ (see also Li and Luo (2012)).
- s-forwards: as proposed in Dowd (2003), Blake et al. (2006) and by the Life and Longevity Markets Association,[17] these are forward contracts on the probability of survival of a cohort from inception at time t_0 to maturity.
- e-forwards: as discussed in Denuit (2009), period life expectancy is a natural index to use for summarising the evolution of mortality rates in a population, and therefore we consider the potential of a forward market in period life expectancy (which we refer to as "e-forwards" from the demographic symbol for period life expectancy) at age x in future year t for hedging purposes.

In each of these cases, we assume that the reference population for the index is the national population used to estimate the APC model underpinning the forward mortality model. Hence, the value of the mortality index at time τ is calculated as:[18]

$$\text{q-forward:} \quad \mathcal{Q}_{x,t}(\tau) = 1 - \exp\left(-\nu^{\mathbb{Q}}_{x,t}(\tau)\right) \tag{12.22}$$

$$\text{s-forward:} \quad \mathcal{S}_{x,t_0,t}(\tau) = {}_{\tau-t_0}p_{x,t_0} \times {}_{t-\tau}P^{\mathbb{Q}}_{x+\tau-t_0,\tau} \tag{12.23}$$

$$\text{e-forward:} \quad \mathcal{E}_{x,t}(\tau) = 0.5 + \sum_{u=0}^{\infty} \exp\left(-\sum_{v=0}^{u} \nu^{\mathbb{Q}}_{x+v,t}(\tau)\right) \tag{12.24}$$

Thus, we can see that these mortality measures are qualitatively different from each other, and range from q-forwards which are very simple securities based on only one

[17] http://www.llma.org/

[18] Note that the s-forward is defined on a reference cohort aged x at the inception data, $t_0 \leq \tau$, and therefore the survivorship of this cohort is a product of the observed survivorship from t_0 to τ, given by ${}_{\tau-t_0}p_{x,t_0}$, and the anticipated survivorship from τ to maturity, t, given by ${}_{t-\tau}P^{\mathbb{Q}}_{x+\tau-t_0,\tau}$. For the purposes of this study, we shall assume that $t_0 = \tau$.

forward mortality rate, to more complex securities which look at forward mortality rates across a number of different ages and years.

For a general forward contract, linked to mortality index $\mathcal{I}_{x,t}$, the forward price specified by the contract must be equal to the time τ value of the mortality measure, i.e., $\mathcal{I}_{x,t}(\tau)$, in order for the contract to have zero value at inception. We assume that the buyer of the contract will receive a floating payment and pay a fixed amount at time t. Hence, the value of the forward contract at time $\tau + 1$ will be

$$B(\tau + 1, t) \left[\mathcal{I}_{x,t}(\tau + 1) - \mathcal{I}_{x,t}(\tau)\right]$$

and, therefore, we are interested in the distribution of the change in the index of mortality over time

$$[\mathcal{I}_{x,t}(\tau + 1)|\mathcal{F}_\tau] - \mathcal{I}_{x,t}(\tau)$$

Although longevity risk is a long-term risk which will materialise over a number of decades, it is likely that longevity-linked securities will need to be considerably shorter-term contracts in order to appeal to speculators. Hence, we only consider forward contracts with maturities of 5, 10 and 15 years, i.e. $t = 5, 10, 15$. Specifically, we investigate the time $\tau + 1$ values of the following forward contracts entered into at time τ:

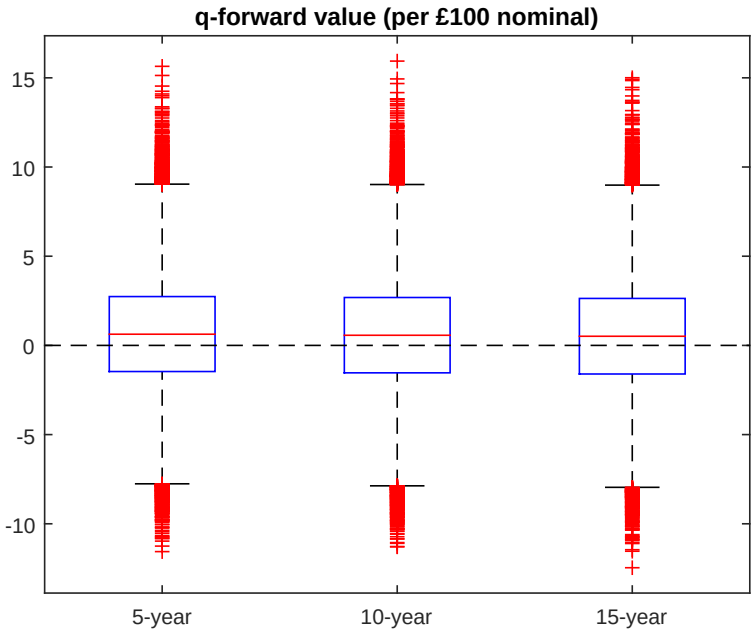
- a q-forward at age 65 and maturity $\tau + t$, i.e., $\mathcal{Q}_{65,\tau+t}$;
- an s-forward with maturity date $\tau + t$, specified on a reference cohort aged 65 at time τ, i.e., $\mathcal{S}_{65,\tau,\tau+t}$; and
- an e-forward at age 65 with maturity $\tau + t$, i.e., $\mathcal{E}_{65,\tau+t}$.

Boxplots showing the time $\tau + 1$ distribution of these forward contracts per £100 of nominal value are shown in Figure 12.9.

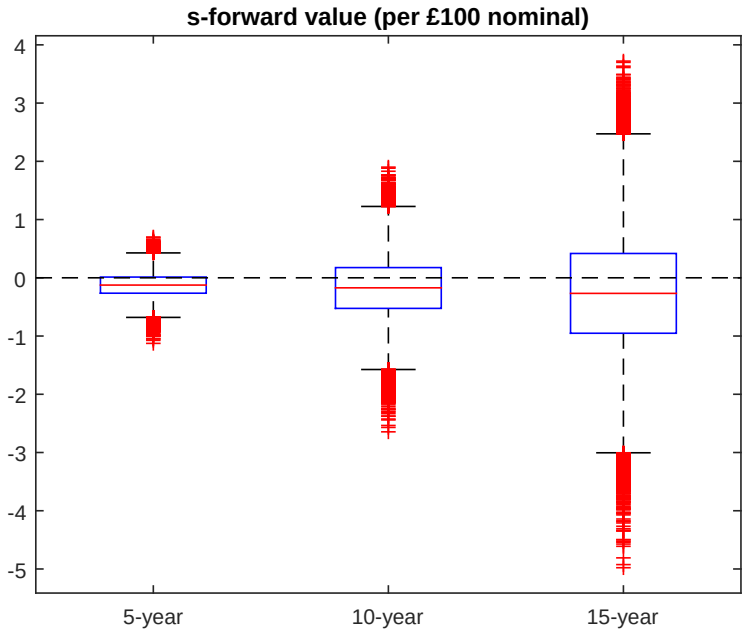
As discussed earlier in the context of annuity values, we note that

$$\mathbb{E}^{\mathbb{P}}{}_\tau \left[\mathcal{I}(\tau + 1)|\mathcal{F}_\tau\right] - \mathcal{I}(\tau) \neq 0$$

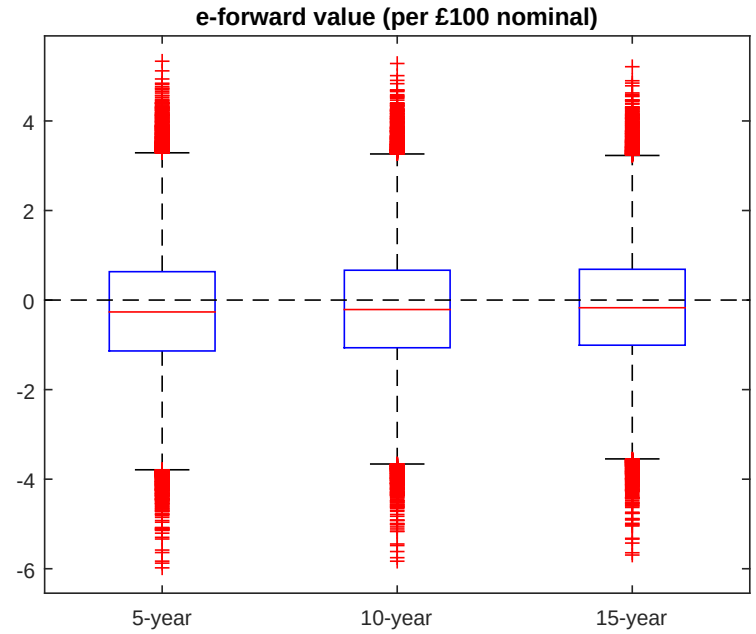
i.e., the expected value of the forward contract at time $\tau + 1$ is not equal to zero, the value at inception. This is, again, due to the prices of securities in the market-consistent measure not being martingales under one-year updates of the forward mortality surface in the real-world measure. Hence, there will be an expected return from trading in



(A) q-Forward



(B) s-Forward



(C) e-Forward

FIGURE 12.9: Boxplots showing the distribution of the values of different longevity-linked securities at time $\tau + 1$

longevity-linked forwards, which arises for the same reasons as the expected release of reserves in annuities, as discussed in Section 12.3.1.

We also see that for q-forwards and e-forwards, the one-year riskiness of the contract does not change significantly with its term. In contrast, the riskiness of an s-forward increases rapidly with the term of the contract. The reason for this is that the nominal value of the mortality index for q-forwards and e-forwards (probability of dying and period life expectancy) does not change much with term, whilst that of the s-forward (survivorship of a cohort) decreases rapidly. This means that longer term q-forward and e-forward contracts could, potentially, be written, with the risk in them managed by annually rebalancing the portfolio. However, this may be more difficult for long-term s-forward contracts and it may be difficult to attract speculators to trade (and hence create liquidity) in the longer-term contracts.

Figure 12.9 also shows that the q-forward contracts are significantly riskier per £100 nominal than the alternatives. This is because the nominal value of the mortality measure is relatively small,[19] and hence the value of the contract is proportionally more affected by new information. In addition, the q-forward is specified on mortality rates at one specific age and time (rather than across a range of ages and years, as in the case of the s-forward and e-forward) which is likely to be more volatile.

When writing forward contracts, it is also necessary to consider the amount required in order to collateralise the contract (which is highly desirable to reduce credit risk in the contract). If the contracts were exchange traded, this amount would also form the basis of the margin account. Assuming the collateral account is readjusted on an annual basis, a sensible method of determining the amount required in the account would be to find the capital needed to protect against a 95% loss on the forward contract, i.e., the economic capital of the contract.[20] The 95% economic capitals for the three ten-year forward contracts per £100 nominal are shown in Figure 12.10.

We see that the 95% economic capital is substantially higher for the q-forward (around 5% to 6% per £100 nominal) than for the s-forward and e-forward. This is consistent with the results shown in Figure 12.9, which indicated that q-forwards over a range of terms were substantially riskier than the alternative contracts. In the other two cases, we

[19] Typically, $q_{x,t}$ will be in the range $[0.005, 0.05]$ for most ages of interest, whilst ${}_{t-t_0}p_{x,t_0}$ will be in the range $[0.1, 0.9]$ and $e_{x,t}$ will be in the range $[10, 30]$.

[20] For Equation 12.15, we see that we are unable to define economic capital ratios for the contracts as they have zero value initially, i.e., $X_0 = 0$.

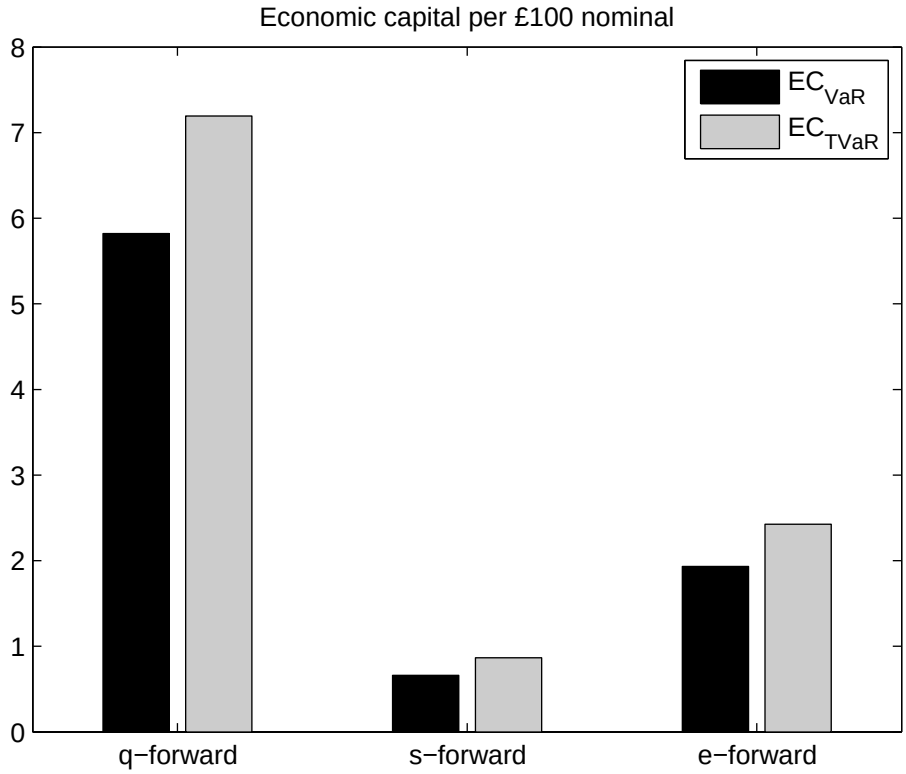


FIGURE 12.10: Economic capital for different longevity-linked securities

find collateral requirements of a few percent of the nominal value. However, for all three contracts, we find that the amount of collateral needed for the contract is fairly typical of other traded forward contracts, such as the standard financial and commodity futures traded on the London International Financial Futures and Options Exchange (LIFFE).

12.3.3.3 Hedging longevity risk

Having measured the longevity risk in annuity values in Section 12.3.3, it is natural to consider how this risk could be managed and reduced. In practice, this can be achieved through reinsurance, securitisation (e.g., Cowley and Cummins (2005)) or natural hedging (e.g., Cox and Lin (2007)). Another method which has been proposed (but not yet widely implemented) is to hedge the longevity risk in a liability portfolio using standardised, tradable longevity-linked securities.[21]

[21] We draw a slight distinction between such a strategy and purchasing a single, customised asset without the intention of rebalancing the hedge in future. Examples of these customised assets include bespoke longevity swaps, as considered in Chapter 10, and highly customised bespoke options on mortality, such as those discussed in Michaelson and Mulholland (2014). However, we feel that this alternative strategy has more in common with a reinsurance policy than truly hedging risk using capital market securities.

To illustrate the potential effectiveness of hedging these illustrative liabilities, we consider using each of the different securities discussed in Section 12.3.3.2 in turn. We adopt a simple mean-variance hedging strategy and select the portfolio whose value at time $\tau+1$ has smallest variance, i.e., we find the hedged portfolio

$$\mathcal{L}^* = \mathcal{L} - \tilde{\theta}\mathcal{I}_{x,t}$$

where $\tilde{\theta}$ is chosen by minimising the variance

$$\begin{aligned}
\tilde{\theta} &= argmin_\theta \mathbb{V}ar_\tau^{\mathbb{P}}\left(\mathcal{L}(\tau+1) - \theta\mathcal{I}_{x,t}(\tau+1)\right) \\
\Rightarrow \tilde{\theta} &= \frac{\mathbb{C}ov_\tau^{\mathbb{P}}\left(\mathcal{L}(\tau+1), \mathcal{I}_{x,t}(\tau+1)\right)}{\mathbb{V}ar_\tau^{\mathbb{P}}\left(\mathcal{I}_{x,t}(\tau+1)\right)} \\
\mathbb{V}ar_\tau\left(\mathcal{L}^*(\tau+1)\right) &= \left(1-\rho_{\mathcal{L},\mathcal{I}}^2\right)\mathbb{V}ar_\tau\left(\mathcal{L}(\tau+1)\right)
\end{aligned}$$

Hence we see that such a strategy depends critically upon the correlation between the liabilities and the hedging instrument, $\rho_{\mathcal{L},\mathcal{I}}$, at time $\tau+1$, with correlations closer to ± 1 giving more effective hedges. The measured correlations for the four securities considered are shown in Table 12.2. Because we wish to minimise the variability of the value of the portfolio at time $\tau+1$, this approach investigates "value" hedging strategies as opposed to "cashflow" hedging strategies, which seek to minimise the uncertainty in the realised cashflows.

Security		q-forward	s-forward	e-forward
Term	5	-93.9%	87.8%	99.5%
	10	-93.9%	89.8%	99.6%
	15	-93.7%	93.9%	99.6%

TABLE 12.2: Correlation between $\mathcal{L}(\tau+1)$ and security values with different terms

As can be seen from Table 12.2, most of the securities being considered give very high correlations with the liabilities. In the case of q-forwards, this correlation is negative, since higher than anticipated reductions in mortality rates have the effect of increasing liability values, but triggering net payments from the buyer to the seller of the q-forward, giving a negative value under the convention adopted in Section 12.3.3.2. This means that a holder of longevity risk will want to receive the floating leg of a q-forward, as opposed to wanting to receive the fixed legs of the other forward contracts.

The high correlations shown in Table 12.2 arise from the same reasons that we observed high correlations between annuity values at different ages in Section 12.3.1. This was because relatively few factors (i.e., the age/period terms in the model, and mainly $\kappa_t^{(1)}$)

drive the changes in mortality rates. These results are therefore model dependent, as cautioned against by Zhu and Bauer (2014). However, we note that the three age/period term model constructed by the general procedure will give more complicated dynamics for mortality (and hence, lower correlations) than most other widely used mortality models, such as the other APC models considered in Chapter 11.

We also note that, for q-forwards and e-forwards, the correlation between the forward contract and the liabilities is roughly independent of the term of the contract. In contrast, the s-forward value becomes more highly correlated with the liability value as the term of the contract increases. This in unsurprising, since longer term s-forward contracts are more exposed to the cumulative effects of longevity risk and will behave more like annuity contracts by their nature. However, as discussed in Section 12.3.3.2 and shown in Figure 12.9, longer term s-forwards are also more risky. This may limit the development of the market in long-term s-forwards which, unfortunately, are amongst the contracts which are most useful for hedging longevity risk.

Figure 12.11 shows the empirical distributions of the value of the unhedged and hedged liabilities (using the three different hedging securities with maturities of ten years) based on 50,000 Monte Carlo simulations. As expected, all the hedging strategies considered appear to substantially reduce the variability of the portfolio value at time $\tau + 1$. This is shown by the economic capital ratios using the VaR and TVaR risk measures (at the 95% level) and the corresponding reductions in risk from the unhedged liability value in Table 12.3.

	ECR_{VaR}	% Reduction of ECR	ECR_{TVaR}	% Reduction of ECR
Unhedged	2.44%	-	3.06%	-
q-forward	0.84%	65%	1.07%	66%
s-forward	1.08%	55%	1.37%	55%
e-forward	0.21%	91%	0.27%	91%

TABLE 12.3: Impact of hedging strategies on longevity risk

It is noticeable from Figure 12.11 and Table 12.3 that the strategy based on an e-forward is significantly more effective at reducing risk than the other two. This is because the values of the period life expectancy at the maturity date is calculated in a similar manner to the calculation of an annuity but over a range of different cohorts, and therefore this security is sensitive to the same risk factors as the annuities we are trying to hedge. In contrast, the q-forward is sensitive to mortality rates at a single selected age, whilst the s-forward considers only a single cohort, and consequently both are poorer at hedging

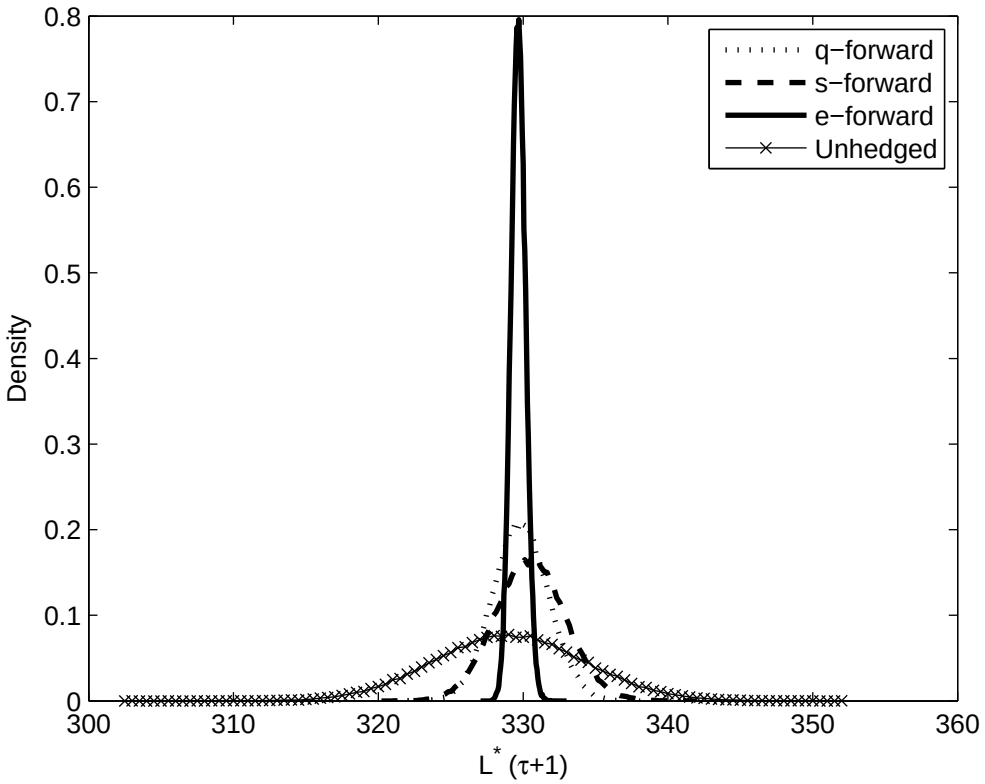


FIGURE 12.11: Empirical distribution of liability values under different hedging strategies

risk.

As can be seen, the reduction in longevity risk with even relatively simple hedging strategies over a one-year period is very high. These are "value hedges", in the sense that the strategy has been chosen to minimise the variance of the total portfolio value, as opposed to "cashflow" hedges that minimise the variability of the net cashflows from the portfolio.[22] Longer term hedges could potentially be achieved by rebalancing the portfolio at least annually to reflect the actual experience of the annuity book. However, such a strategy is dependent upon the existence of a relatively liquid market in the underlying longevity-linked securities.

One potential criticism is that these results are all model dependent. It does not seem likely that the high correlations shown in Table 12.2 could be achieved in practice and, therefore, such large reductions in risk may not be feasible. In particular, the use of relatively simple APC mortality models to underpin the forward mortality framework might be felt to give correlation structures for future mortality rates which are overly

[22]Examples of cashflow hedging solutions for longevity risk include bespoke longevity swaps.

simplistic, and so overstate the effectiveness of any hedging strategy. However, we note that our underlying model for the force of mortality has three age/period terms and a cohort term, making it relatively complex compared with many more commonly used mortality models, and so it is unlikely that using a more complicated model for the short rate would materially affect our results.[23] In addition, the impact of hedging would be lower if the market prices of risk change during the year. However, since the market for longevity risk is just emerging, assuming constant market prices of risk is unavoidable at present, for the reasons discussed in Section 12.2, and, accordingly, all liability and securities values will be model-dependent for the foreseeable future. Furthermore, high correlations between the liabilities and hedging instruments are required in order to recognise the hedge under some accounting standards. Therefore, we argue that reductions in risk, even if they are only mark-to-model, are still beneficial for many purposes.

In addition, the results presented above do not allow for potential basis risk between populations or for idiosyncratic risk in the number of deaths observed in an actual annuity book, and so will overstate the potential effectiveness of hedging strategies which could be obtained in practice. We leave the addition of both of these sources of risk to future work.

12.4 Multi-year risk measurement and the Solvency II risk margin

12.4.1 Projecting the liabilities

In Section 12.3, we considered the possible changes in the values of a portfolio of annuities over a one-year period. However, it should be clear that we can also use the forward mortality rate framework to measure longevity risk in the liabilities over a longer time horizon than just one year. This is especially valuable as longevity risk is a long-term risk which may take years or decades to fully emerge.

[23] We have tested the hedging strategies using the simpler models of the short rate of mortality discussed in Chapter 11 and obtain even higher reductions in risk. In particular, we observed perfect correction between the liabilities and securities, and therefore perfect hedges, when using the Lee-Carter model as the underlying mortality model, since this model only possesses one age/period term and hence only one source of risk.

To do this, we start by extending the definition of the liabilities in Equation 12.17 to allow for multiple years, i.e., the liability at time t is equal to

$$\mathcal{L}(t) = \sum_{x=60}^{80} \left[\left(\sum_{s=\tau+1}^{t} B(\tau, s)_{s-\tau} p_{x,\tau} \right) + B(\tau, t)_{t-\tau} p_{x,\tau} a_{x+t-\tau}(t) \right] \tag{12.25}$$

Similar to Equation 12.17, using this form for the liabilities allows for the impact of benefits paid and interest, and therefore ensures that $\mathcal{L}(t)$ is comparable to $\mathcal{L}(\tau)$.

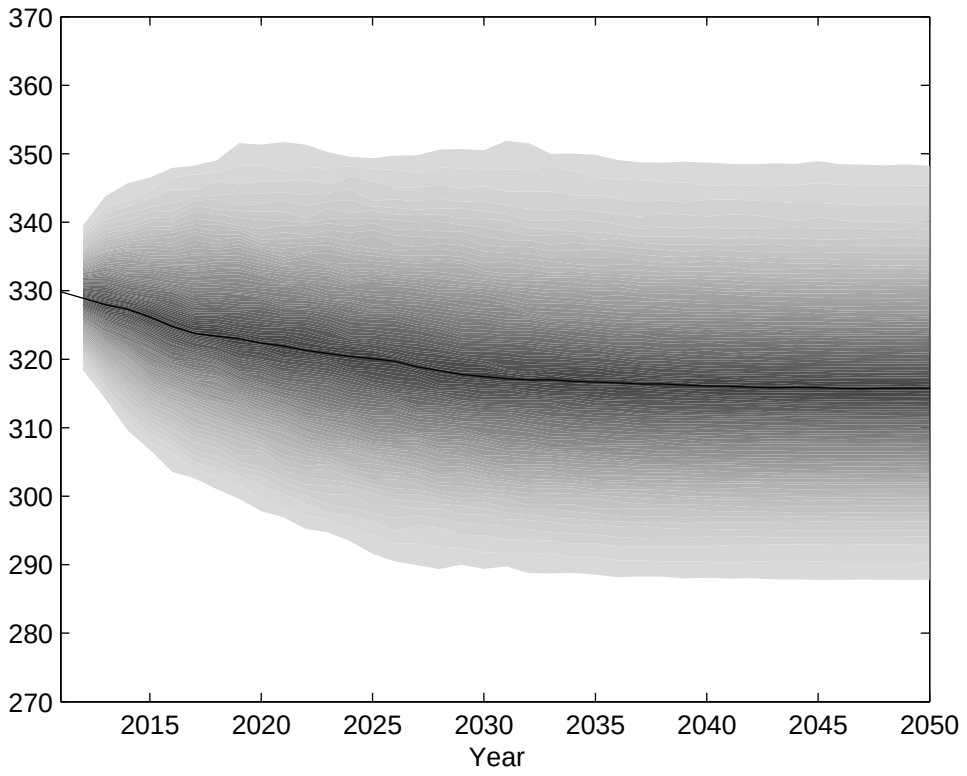


FIGURE 12.12: Distribution of future market-consistent liability values

Projected market-consistent liability values are shown in Figure 12.12. The first thing to note about these is that the variability in the liabilities increases rapidly in the first year, but then grows more slowly over the remaining term of the benefits. This is because changes in the estimation of future mortality rates have a greater impact while the liabilities are relatively immature than when most of the benefits have already run off. This can be considered analogous to the interest-rate risk in a portfolio of bonds, which decreases with time as the bonds mature and the duration of the portfolio decreases.

In addition, we note that median of the liabilities decreases with time, from the initial value of 331.4 (in notional currency units) to 316.4, i.e., a decrease of approximately

4.5% over the lifetime of the liabilities. This is due to the release of reserves over the period, as discussed previously in the single-year context in Section 12.3. This is caused by the market-consistent liability value being greater than a true "best estimate" of the present value of the future benefits, i.e., the real-world liability value.

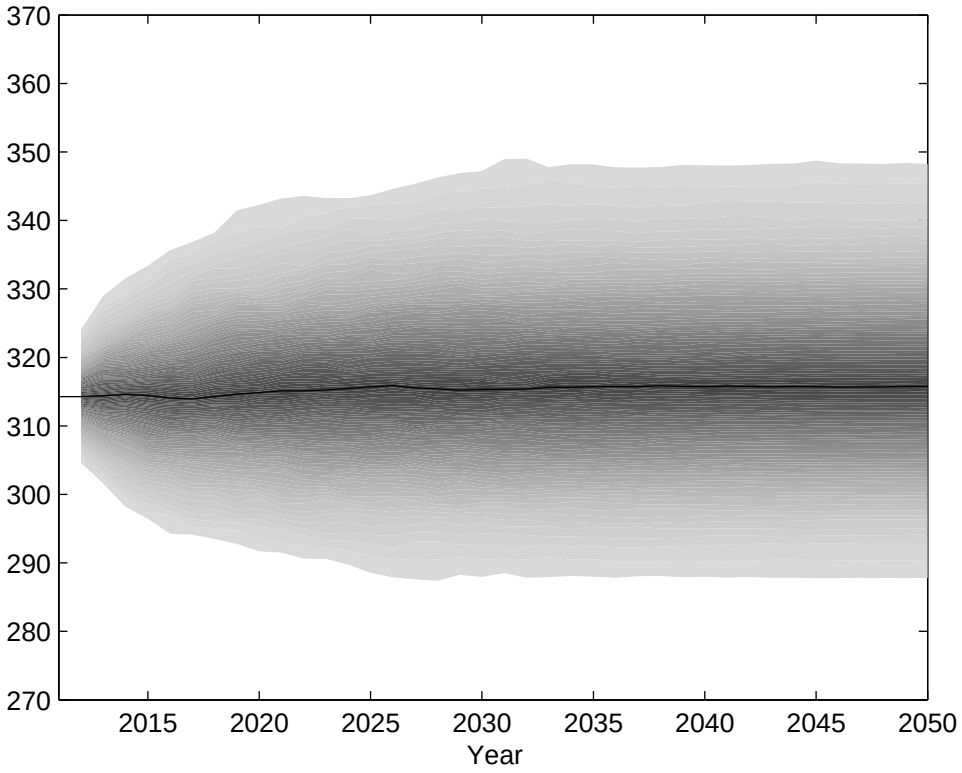


FIGURE 12.13: Distribution of future real-world liability values

Correspondingly, Figure 12.13 shows the projected liabilities if they are valued using the real-world approach, as opposed to the market-consistent approach above. Here, we observe very similar levels of riskiness in the liabilities, but no release of reserves, which is consistent with the results of Section 12.3.3.1 over a single year. Indeed, the distribution of the real-world and market-consistent liabilities ultimately converge to the same distribution, since this is given by the projected benefits paid during run-off, which is determined in the real-world measure. However, since the liabilities are systematically lower in the $\mathbb{P}$-measure compared with the $\mathbb{Q}$-measure, using these liabilities values alone as the technical provisions under Solvency II would be inconsistent with the desire to achieve a market-consistent approach for reserving for life insurance liabilities.

12.4.2 The Solvency II risk margin

To allow for the difference between the real-world and the market-consistent valuation of the liabilities, EIOPA (2014) requires insurers to add a "risk margin" to the real-world liability value as a proxy for the additional cost required to transfer the liabilities to a third party. Specifically, *"The risk margin is a part of technical provisions in order to ensure that the value of technical provisions is equivalent to the amount that insurance and reinsurance undertakings would be expected to require in order to take over and meet the insurance and reinsurance obligations"* (EIOPA (2014, T.P.5.2.)). In order to proxy for this, EIOPA assumes that a reinsurer would require an additional amount equal to the future costs of holding sufficient capital to insure the risk, i.e., the present value of future SCRs. Therefore, the risk margin is defined in EIOPA (2014) as

$$\text{Risk Margin}(\tau) = \text{CoC} \times \sum_{t=\tau}^{\infty} \text{SCR}(t) B(\tau, t) \tag{12.26}$$

where the SCR is defined as in Section 12.3.3.1 and CoC is the cost of capital for the annuity business.[24] To avoid having a circular definition, the SCR is defined as the value at risk of changes in the real-world liability value, not the technical provisions (which would also include the risk margin, and hence depend upon the value of the SCR). In addition, the SCR at time t is a random variable since it is conditional on $\mathcal{F}_t$ and so will depend upon the evolution of mortality rates and the liabilities between time τ and t. This is discussed further in Christiansen and Niemeyer (2014). In order to calculate the risk margin, we use the modified definition

$$\text{Risk Margin}(\tau) = \text{CoC} \times \sum_{t=\tau}^{\infty} \mathbb{E}^{\mathbb{P}}{}_{\tau}\text{SCR}(t) B(\tau, t) \tag{12.27}$$

where we have taken expectations of the SCR conditional on the initial information at time τ.

However, calculating the risk margin is problematic as "nested" simulations (i.e., simulations within simulations) are required, as discussed in Bauer et al. (2012). This is because:

1. Monte Carlo simulations are required to project the liabilities from time τ to time t stochastically. In order to obtain a fair sample of the distribution of the liabilities at t, a large number (say, N) of Monte Carlo simulations are required for this.

[24] In line with EIOPA (2014, TP.5.21), we use a cost of capital of 6% p.a..

2. For each projected liability at time t, Monte Carlo simulations are required to calculate SCR(t), which requires the the stochastic development of the forward mortality surface from time t to time $t+1$. Since the SCR is the value at risk at a very high significance level, a large number (say, $M \geq N$) of Monte Carlo simulations are also needed.

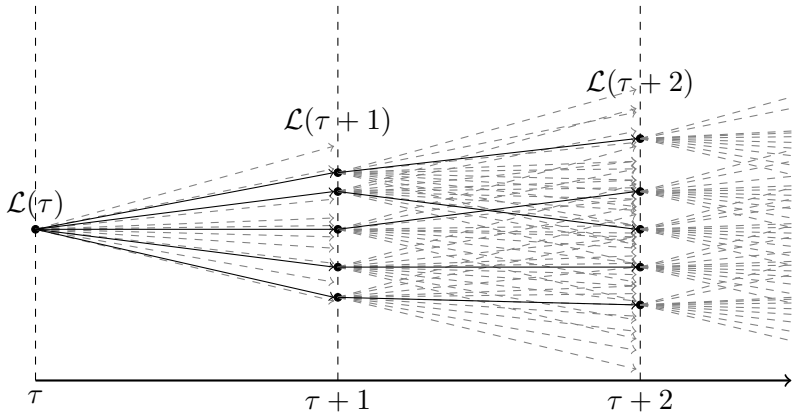


FIGURE 12.14: "Nested simulations" approach for calculating the risk margin, $N = 5$ simulations used to project the liabilities and $M = 10$ simulations (dashed) to calculate the SCR at each future time for each liability value

This is illustrated in Figure 12.14, showing $N = 5$ simulations for projecting the liabilities and $M = 10$ simulations in order to calculate the one-year update of the liabilities at each future time in order to calculate the SCR. Using this "nested" approach with $N = 1,000$ and $M = 20,000$, we calculate a risk margin of 10.3 notional currency units, equivalent to 3.3% of the real-world liability values. As shown in Table 12.4, this would give total technical provisions of 324.5 (in notional currency units), compared with 331.4 if the technical provisions are calculated using the market-consistent valuation of the liabilities and so would not fully compensate for the difference between the real-world and our illustrative market-consistent measure.

Approach	Liability Value	Risk Margin	Technical Provisions	SCR(τ)	Total Liabilities
Market-consistent	331.4	-	331.4	12.8	344.2
Real-world	314.2	10.3	324.5	12.6	336.4

TABLE 12.4: Technical provisions and SCRs using different approaches

In addition, Figure 12.15 shows the projected SCRs in future years as the liabilities are run off. We see from this that the SCR is expected to decrease rapidly as the benefits are run-off and so decrease in riskiness.

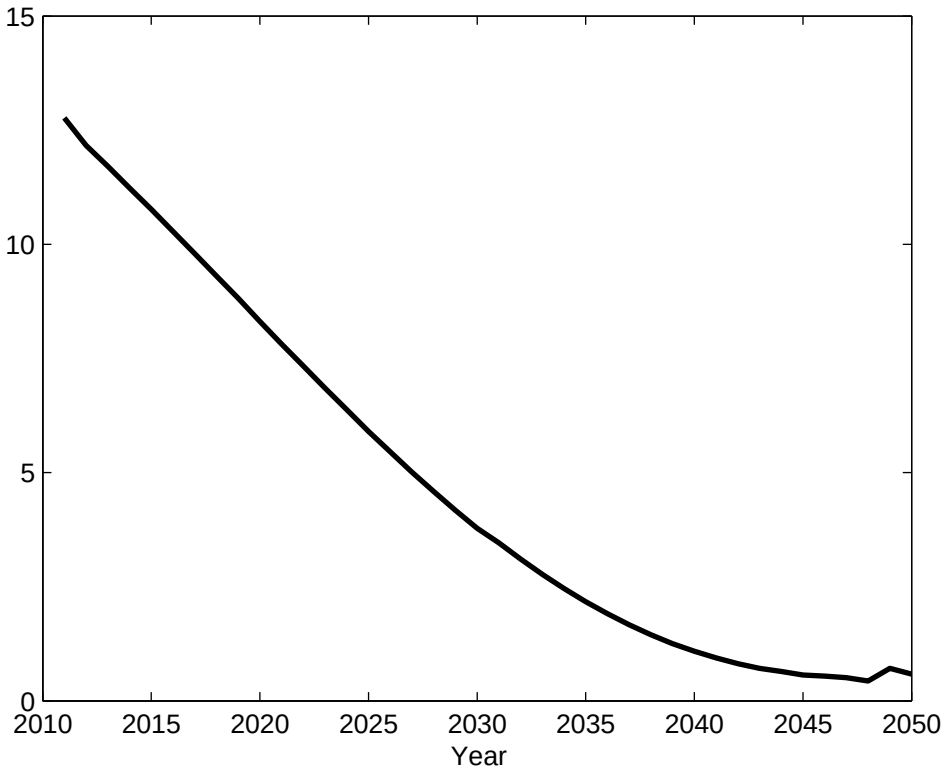


FIGURE 12.15: Future SCR values calculated using nested simulations

However, this nested approach is computationally intensive, as it requires $N \times M$ Monte Carlo simulations. To calculate the risk margin above took a several days of computing time on a single desktop computer. Although using more powerful computers running in parallel could, potentially, reduce this time, the calculation of the risk margin for an insurer would be considerably more complicated, since more risks, other than longevity risk, would need to be included.

12.4.3 Approximate calculation of the risk margin

Since the full calculation of the risk margin is computationally intensive, there have been a number of different methods suggested in order to calculate it approximately by simplifying the calculation. These have, broadly speaking, taken two approaches:

1. Projecting the liabilities from time τ to time t stochastically, but then approximating the calculation of $\text{SCR}(t)$. This reduces the computational burden from $N \times M$ to N. Examples of these techniques are discussed in Section 12.4.3.1.

2. Projecting the liabilities from time τ to time t deterministically, but then calculating SCR(t) exactly. This reduces the computational burden from $N \times M$ to M. Examples of these techniques are discussed in Section 12.4.3.2.

One of the simplest practical methods for simplifying the calculation of the risk margin was proposed in EIOPA (2014, TP.5.60) and uses both of these approaches simultaneously to calculate the risk margin deterministically, not based on any Monte Carlo simulations. This approach calculates the risk margin using the modified duration of the liabilities

$$\text{Risk Margin} = B(\tau, \tau+1) \times \text{CoC} \times \text{Dur}_\tau \times \text{SCR}(\tau)$$

This "duration" approach avoids the need either to do stochastic simulations to project the liabilities, or for additional estimates of the SCR in future years (just an initial value at time τ). It is therefore unlikely to fully capture the uncertainty in the future liabilities and so will provide a relatively crude estimate of the capital required. Using this technique, we estimate a risk margin of 2.5% of the liabilities, based on a duration of 10.5 years and the SCR(τ) from the forward mortality framework. This is significantly below the value obtained from using nested simulations, and indicates that the duration approach may understate the technical provisions if used for the calculation of the risk margin.

12.4.3.1 Approximating the SCR

A number of techniques are available to approximate the SCR at time t without the need to estimate it via Monte Carlo simulations and, therefore, reduce the calculation burden of computing the risk margin. This is illustrated in Figure 12.16, showing the $N = 5$ simulations required to project the liability values, but no nested simulations required to calculate the SCR for each of them.

The first of these techniques was proposed in EIOPA (2014, SCR.7.29) and calculates the SCR for each simulation using the standard model, i.e., by stressing the mortality rates by 20% in each simulation and for each future time to calculate liability values. However, this "standard model" approach suffers from the same disadvantages as were discussed in Börger (2010) and Section 12.3.3 for calculating the initial SCR at time τ. Most importantly, it will systematically result in a large over estimate of the actual capital requirement. However, we can modify this approach by rescaling the standard

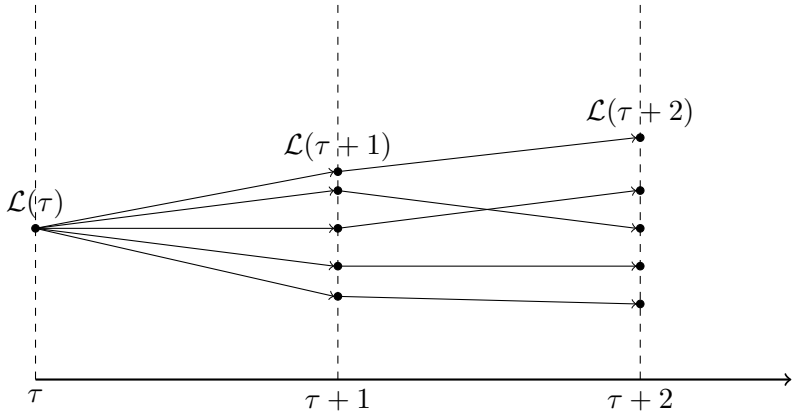


FIGURE 12.16: Approximate approach for calculating the risk margin, using $N = 5$ simulations to project the liabilities but approximating the SCR at each future time for each liability value

model SCR at time t, i.e.,

$$\begin{aligned}\text{SCR}(t) &\approx \text{SCR}_{StandardModel}(t) \times \frac{\text{SCR}_{ForwardRates}(\tau)}{\text{SCR}_{StandardModel}(\tau)} \\ &= \text{SCR}_{StandardModel}(t) \times \frac{4.0\%}{10.0\%} \quad \text{from Section 12.3.3.1}\end{aligned}$$

so as not to systematically overestimate the future values of the SCR.

When we do this, we find a risk margin equal to 5.3% of the best-estimate value of the liabilities using $N = 1,000$, which is higher that that found using the nested simulations approach discussed above. Figure 12.17 shows the projected $\mathbb{E}^{\mathbb{P}}{}_{\tau}$ SCR(t) values using this modified "standard model" approach. This shows that the standard model approach results in an unusual pattern as the liabilities are run off, with the SCR decreasing more slowly at first than in Figure 12.15 using the nested simulations, and then falling rapidly after around 30 years. We regard this as highly unusual, especially considering we would expect the uncertainty in the liabilities to decrease rapidly as they mature. Therefore, we believe that the standard model approach does not provide a good approximation for the full calculation of the risk margin using nested simulations.

A second approach, discussed in EIOPA (2014, TP.5.52), assumes that SCR(t) is proportional to the prospective liability value, i.e., $\text{SCR}(t) = \text{SCR}(\tau) \times \frac{\tilde{\mathcal{L}}(t)}{\tilde{\mathcal{L}}(\tau)}$.[25] Using this technique and $N = 1,000$, we calculate a value for the risk margin of 4.0% of the best-estimate liability value, which is higher than that given by the nested simulations. Values of the projected $\mathbb{E}^{\mathbb{P}}{}_{\tau}$ SCR(t) using this "proportional" approach (which takes the SCR

[25] This uses the prospective liabilities, $\tilde{\mathcal{L}}(t)$, as opposed to the liabilities $\mathcal{L}(t)$ defined in Equation 12.25, i.e., the value of the benefits paid beyond time t, discounted to t.

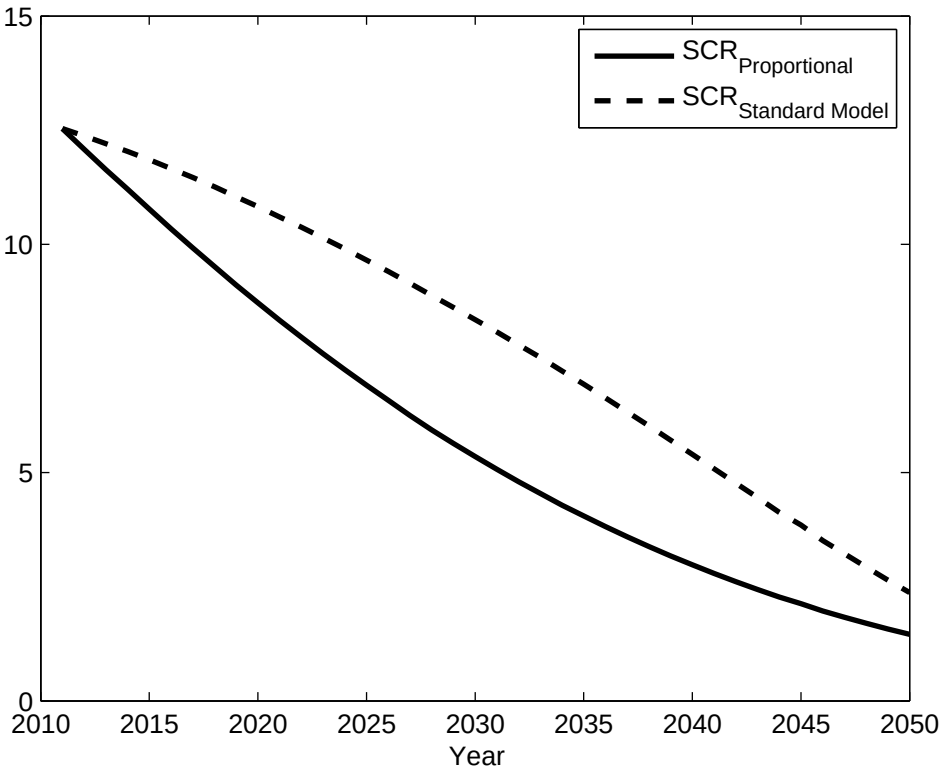


FIGURE 12.17: Projected SCRs using the standard model and proportional approaches

calculated at time τ from the forward mortality rate framework) are also shown in Figure 12.17.

The pattern of future SCRs calculated using the proportional approach is more similar to that shown in Figure 12.15 as the liabilities mature. However, the liabilities at time $t+1$, conditional on time t, will not have the same distribution as the liabilities at time $\tau+1$, conditional on time τ, due to their increasing maturity. This has the potential to distort the estimate of the SCR, and, as the level of longevity risk in more mature annuity portfolios will generally be lower than less mature portfolios, may bias the SCR upwards. However, for the relatively simple illustrative annuity portfolio used in this study, this effect does not appear to be significant and the proportional approach gives a reasonable approximation to the full nested simulations approach.

Both of the approaches discussed above calculate SCR(t) as a relatively simple function of $\mathcal{L}(t)$, the liability value at time t. More complicated functions could also be used to estimate SCR(t), for instance, using the techniques of Denuit (2008) in the context of using the Lee-Carter model as the underlying mortality model, or through the use of

extreme value theory, as mentioned in passing by Bauer et al. (2009). However, these approaches have been developed in relatively simple and specific contexts, and so are unlikely to be feasible for complex liabilities or when using more sophisticated mortality models.

A conceptually similar approach, proposed in Bauer et al. (2009) and based on techniques that are popular in option pricing, is the use of least-squares Monte Carlo methods. This approach uses a number of deterministic scenarios to regress the SCR at time τ as a function of the underlying latent variables of the model (i.e., the period and cohort functions $\boldsymbol{\kappa}_\tau$ and $\gamma_{\tau-x}$). This approach is also conceptually similar to those suggested in Cairns (2011) and Dowd et al. (2011a).

However, complicated mortality models (especially those with cohort parameters) will have a large number of latent variables, e.g., the GP model using in this study has 54 latent variables corresponding to the three period functions and the 51 cohort parameters for years of birth with members which are currently alive. Therefore, it is unclear how practical least squares Monte Carlo methods are for more complicated annuity portfolios and sophisticated mortality models. Least squares Monte Carlo methods are also most suitable for processes whose distributions do not change significantly with time, and therefore may not be appropriate for modelling liabilities in run-off.

12.4.3.2 Approximating the liabilities

A fundamentally different approach is to calculate the SCR at time t accurately using Monte Carlo simulations, but to use only a reduced number of scenarios to model the evolution of the liabilities from τ to t. Börger (2010) and Stevens et al. (2010) suggest using the best estimate (i.e., median) scenario to project the liability value. To do this, we calculate $\mathcal{L}_{Med}(t)$ deterministically, but then use Monte Carlo simulations to project one year ahead and estimate the distribution of $\mathcal{L}(t+1)|\mathcal{L}_{Med}(t)$. From this distribution, $\text{SCR}_{Med}(t)$ can be estimated as

$$\text{SCR}_{Med}(t) = \text{EC}_{VaR}(\mathcal{L}(t+1)|\mathcal{L}_{Med}(t); 99.5\%)$$

This "median" approach is illustrated in Figure 12.18, showing $M = 10$ simulations in grey to calculate the SCR at each future time.

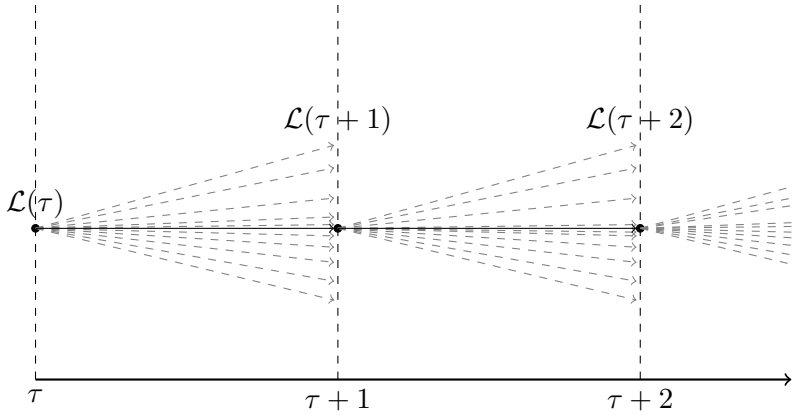


FIGURE 12.18: "Median" approach for calculating the risk margin, using $M = 10$ simulations to estimate the SCR for the median liability value at each future time

The median approach captures the impact of the increasing maturity of the liabilities on the distribution of the one-year projection of the liabilities and hence the estimation of the SCR. Using the median approach with $M = 20,000$, we estimate a risk margin of 3.5% of the real-world value of the liabilities, which is not very different from that calculated using nested simulations.

One potential criticism of this approach is that it does not calculate the SCR for scenarios where the liabilities are already significantly higher or lower than the median estimate at time t. For instance, although at time t, we consider highly adverse scenarios for how mortality might evolve to $t+1$ in the calculation of $\text{SCR}(t)$, we only use the median scenario to calculate $\mathcal{L}(t+1)$ and, hence, ignore the potential for adverse experience to actually be realised. As the SCR is a phenomenon relating to the tail of the distribution of the liabilities, it may have a different distribution if the starting liabilities at time t are already in a stressed scenario compared with the best estimate scenario.[26]

However, this approach can be extended to deal with this criticism. We do this by recognising that the median scenario is just one representative scenario (or "model point") of the distribution of liabilities at time t. As we have to find the distribution of the liabilities at time t in order to calculate $\text{SCR}(t-1)$, we could instead take multiple model points (say, p) from this distribution. For each of these representative scenarios, we would then calculate $\text{SCR}(t)$, with the estimated total SCR at time t being a probability-weighted average of the estimates for each model point. This approach is illustrated in Figure 12.19, using $p = 3$ model points and $M = 10$ simulations to calculate the SCR for each

[26] For example, we may believe that our liabilities will behave differently if we have already observed rapid reductions in mortality rates compared to a best estimate scenario.

model point.

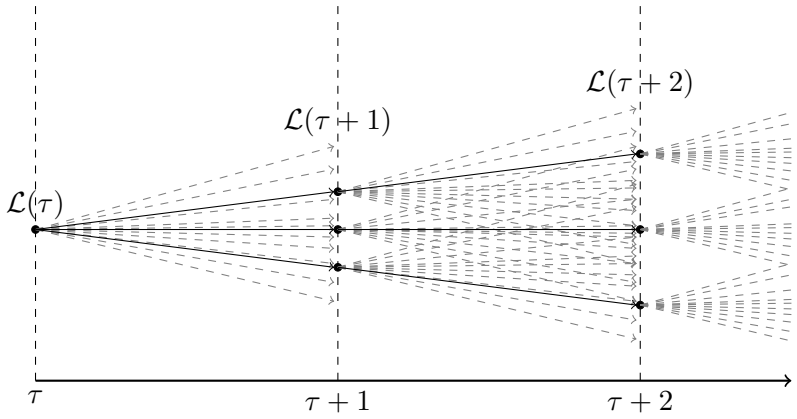


FIGURE 12.19: "Model point" approach for calculating the risk margin, using $p = 3$ model points and $M = 10$ simulations to estimate the SCR for each model point at each future time

Hence, the model point approach is able to investigate the behaviour of the SCR under adverse scenarios for the evolution of the liabilities. This may be important for complicated benefit structures, whose distribution could, potentially, be strongly path-dependent. This "model point" approach requires $\sim p \times M$ simulations, which, although still computationally intensive, is a significant reduction from the $N \times M$ simulations required for the nested approach.

The choice of model points is left to the model user. For illustrative purposes, in our calculations, we have chosen model points at regular quantiles of the liability distribution, namely p model points at the $\frac{1}{2}\frac{100}{p}^{th}$, $\frac{3}{2}\frac{100}{p}^{th}$, ... $\frac{2p-1}{2}\frac{100}{p}^{th}$ percentiles of the distribution, which are all given the same weight. However, other choices might be more appropriate if a greater number of model points in the tails of the distribution is desired, although the weights would have to be adjusted appropriately.

Nevertheless, there is, still a trade-off between the number of model points and the number of simulations to calculate the SCR at each model point for a fixed computational budget. Using more model points, therefore, means reducing the number of simulations used to calculate the SCR for each, in order to keep the same total number of simulations. To illustrate this, we calculate the SCR during run off and the risk margin for different values of p with fixed $p \times M = 20,000$.[27]

[27] Note that the case $p = 1$ corresponds to the median approach discussed above.

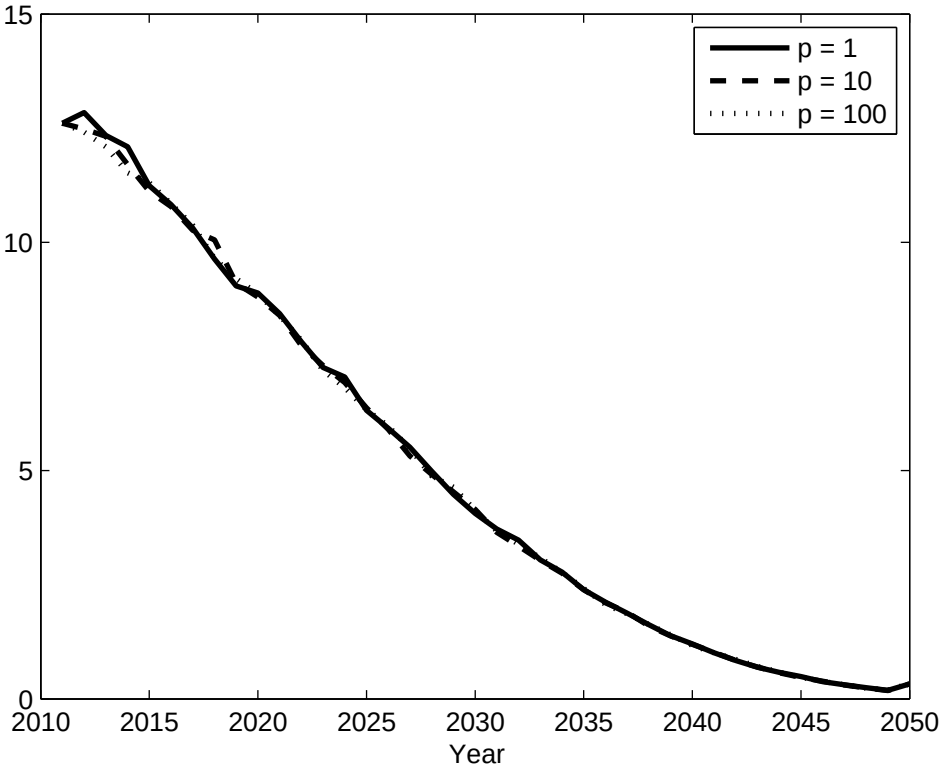


FIGURE 12.20: Projected SCRs for different numbers of model points

Figure 12.20 shows the estimated SCRs at different times and different numbers of model points. As can be seen, the number of model points does not appear to significantly affect the projected values of the SCR, and therefore will give similar estimates of the risk margin. The is because the liability values are driven chiefly by the period parameters, $\boldsymbol{\kappa}_t$, which are projected using a random walk with drift. This means that the period functions are not path-dependent, e.g., observing larger than anticipated changes in $\boldsymbol{\kappa}_t$ between τ and t merely changes the starting point for where we expect the process to go in future, but does not cause us to revise our expectation of the drift of the process. Allowing for experience to feed through into an adjusted assumption for the drift of the process (i.e., allowing for recalibration risk) in the forward mortality framework would change this assumption and, potentially, our results. However, the structure of our illustrative liabilities is relatively simple, which may limit the extent to which SCR(t) would be affected by adverse experience between τ and t. There might be more complicated situations where a greater number of model points are needed in order to capture the behaviour of the SCR in the tail of the liability distribution, such as if the liabilities included longevity-linked options, such as guaranteed annuity rates.[28]

[28] However, as discussed in Chapter 11, the forward mortality rate framework cannot currently be used to value options on measures of mortality. We leave the modelling of longevity-linked options to future work.

12.4.3.3 Comparing the approaches

Values of the SCR and risk margin at time τ (as percentages of the time τ liability value) are shown in Table 12.5 for the various methods considered above for calculating the risk margin. As can be seen, the standard model approach overestimates the amount of capital required, both compared with the approach based on nested simulations and the other approximate approaches discussed. Thus, an insurer using the standard model approach may experience a lower return on capital from their annuity book and find writing annuities less profitable than competitors adopting a more sophisticated approach. In contrast, we find that the duration approach underestimates the risk margin and, hence, the total capital required, which might prompt further investigation from the regulator regarding the capital adequacy of an insurer using this approach. The other methods find broadly comparable amounts of risk capital in order to support the annuity book. However, our results are based on a very simple, illustrative annuity book and therefore may not be directly applicable for the more realistic annuity books.

Approach	SCR(τ)	Risk Margin	Total
Nested	4.0%	3.3%	7.3%
Duration	4.0%	2.5%	6.5%
Standard model	4.0%	5.3%	9.3%
Proportional	4.0%	4.0%	8.0%
Median	4.0%	3.5%	7.5%
Model point ($p = 10$)	4.0%	3.5%	7.5%

TABLE 12.5: SCRs and risk margins using different approaches

In practice, any measurement of the SCR and risk margin would also need to take into account other risks, such as uncertain investment returns, interest rates, inflation, policyholder behaviour and operational risks, all of which might add substantially to these requirements. It is important that any model used for longevity risk can be integrated into the wider framework of measuring and managing the full range of risks faced by a life insurer. We believe that approaches which provide greater detail about the potential evolution of the liabilities, such as the model point approach, can do this more effectively and, hence, provide a more holistic approach to risk management within the annuity book, than some of the other simpler approaches discussed above.

12.5 Conclusions

In Chapter 11, we defined a static forward surface of mortality for the purpose of valuing longevity-linked liabilities and securities. In this study, we extend this framework by investigating the dynamics of the forward mortality surface to show how these values might change over time. This involves understanding the processes we use to project the underlying parameters in the mortality model and how these update to reflect new information. In particular, an understanding of how the cohort parameters in the model update in response to new information is critical in measuring the dynamics of the forward mortality surface. We use this understanding to show that forward mortality rates are martingales in both the real-world and market-consistent measures, and are, therefore, "self-consistent" in the terminology of Zhu and Bauer (2011b).

We then apply this dynamic framework to investigate some of the most important current issues in the measurement and management of longevity risk. In particular, we demonstrate how the forward mortality framework could be used as an internal model as part of the Solvency II regulations being implemented across the EU. We also compare it with the standard model proposed in EIOPA (2014), which we have demonstrated significantly overstates the amount of capital an insurer would need to hold in respect of longevity risk. We also investigate the calculation of the Solvency II risk margin and compare a variety of approaches for simplifying this. In addition, we use the forward mortality framework to investigate the effectiveness of longevity-linked securities in hedging longevity risk in an annuity portfolio, and find that relatively simple hedging strategies can significantly mitigate the longevity risk in a set of illustrative liabilities.

However, the forward mortality framework described here and in Chapter 11 contains some notable omissions, namely that it cannot currently allow for revisions to the trend rate of mortality improvement (recalibration risk in the terminology of Cairns (2013)), does not allow for potential basis risk between populations and cannot be used to value options on mortality rates and other instruments with non-linear longevity-linked payoffs. We leave each of these problems for future work, but are confident that they are solvable.

In Chapter 11, we stated our belief that the forward mortality rates are the way forward in answer to the question posed in Norberg (2010). This study reaffirms this conclusion and demonstrates the many practical uses a forward mortality framework can have in completing the framework for measuring and managing longevity risk.

12.A Self consistency

In Section 12.2, we discussed the self-consistency property of Zhu and Bauer (2011b) and argued that $\mathbb{P}$-measure forward mortality rates should be self-consistent in the real-world measure and $\mathbb{Q}$-measure forward mortality rates should be self-consistent in the market-consistent measure since they are defined as conditional expectations. However, it is helpful to confirm this explicitly in order to ensure that there are no inconsistencies in the modelling framework. This was done for age/period models of the short rate in Section 12.2.1, where the time series process updating the period parameters was relatively simple. In this Appendix, we first verify the martingale property for models that include a cohort term and then verify that forward mortality rates are self-consistent in the market consistent $\mathbb{Q}$-measure.

12.A.1 Self consistency of the cohort parameters

For simplicity, consider a model of the short mortality rate with no age/period terms, i.e.,

$$\ln \mu_{x,t} = \alpha_x + \gamma_{t-x}$$

In this case

$$\nu^{\mathbb{P}}_{x,t}(\tau) = \exp\left(\alpha_x + M(t-x,\tau) + \frac{1}{2}V(t-x,\tau)\right)$$

and trivially therefore

$$\nu^{\mathbb{P}}_{x,t}(\tau+1) = \exp\left(\alpha_x + M(t-x,\tau+1) + \frac{1}{2}V(t-x,\tau+1)\right)$$

First, we observe that

$$V(y,\tau+1) = V(y-1,\tau) \tag{12.28}$$

from the definition of the variance function in Equation 11.21. The, using Equation 12.28 and dropping the superscript $\mathbb{P}$ (since all expectations and variances are in the real-world measure), we see that self-consistency implies

$$\begin{aligned}\exp\left(\alpha_x + M(t-x,\tau) + \frac{1}{2}V(t-x,\tau)\right) &= \mathbb{E}_\tau\ \exp\left(\alpha_x + M(t-x,\tau+1) + \frac{1}{2}V(t-x,\tau+1)\right)\\ &= \exp\left(\alpha_x + \mathbb{E}_\tau M(t-x,\tau+1) + \frac{1}{2}\left(\mathbb{V}ar_\tau(M(t-x,\tau+1)) + V(t-x-1,\tau)\right)\right)\end{aligned}$$

Therefore, we require

$$\mathbb{E}_\tau M(y,\tau+1) = M(y,\tau) \tag{12.29}$$

$$\mathbb{V}ar_\tau\left(M(y,\tau+1)\right) = V(y,\tau) - V(y-1,\tau) \tag{12.30}$$

It is important to note that these are direct consequences of the laws of conditional expectation and variance, i.e., Equations 12.29 and 12.30 can be rewritten as

$$\mathbb{E}_\tau \mathbb{E}_{\tau+1}\gamma_y = \mathbb{E}_\tau \gamma_y$$

$$\mathbb{V}ar_\tau\left(\mathbb{E}_{\tau+1}\gamma_y\right) + \mathbb{E}_\tau \mathbb{V}ar_{\tau+1}(\gamma_y) = \mathbb{V}ar_\tau(\gamma_y)$$

and therefore that the following is merely a check on whether the Bayesian process underpinning the cohort parameter is internally consistent.

For simplicity, we assume that we have chosen a set of identifiability constraints such that $\beta = 0$. From Chapter 6, we have the following recursive relationships which define the mean and variance functions (and which were solved to give the closed forms of $M(y,\tau)$ and $V(y,\tau)$ in Equations 11.20 and 11.21)

$$M(y,t) = \underline{\gamma}_y(t) + (1 - D_{t-y})\rho M(y-1,t) \tag{12.31}$$

$$V(y,t) = (1 - D_{t-y})\sigma^2 + (1 - D_{t-y})^2\rho^2 V(y-1,t)) \tag{12.32}$$

Starting with Equation 12.29

$$\begin{aligned}\mathbb{E}_\tau M(y,\tau+1) &= \mathbb{E}_\tau \sum_{s=0}^{\infty}\left[\prod_{r=0}^{s-1}(1 - D_{\tau+1-y+r})\right]\rho^s \underline{\gamma}_{y-s}(\tau+1) \\ &= \sum_{s=0}^{\infty} Þ_{\tau+1-y,s}\rho^s \mathbb{E}_\tau \underline{\gamma}_{y-s}(\tau+1)\end{aligned}$$

where we have defined

$$Þ_{\tau-y,s} = \prod_{r=0}^{s-1}(1 - D_{\tau-y+r})$$

and $Þ_{\tau-y,0} = 1$ by definition, as per Chapter 6. From this definition, we note the following

$$Þ_{\tau-y,s+1} = (1 - D_{\tau-y+s})Þ_{\tau-y,s}$$

$$Þ_{\tau-y+1,s} = \frac{(1 - D_{\tau-y+s})}{(1 - D_{\tau-y})}Þ_{\tau-y,s}$$

From Equation 12.7 we have

$$\mathbb{E}_\tau \gamma_y^{\tau+1-y} = \rho M(y-1,\tau)$$
$$\mathbb{V}ar_\tau(\gamma_y^{\tau+1-y}) = \rho^2 V(y-1,\tau) + \frac{\sigma^2}{d_{\tau+1-y}}$$

Using this with Equation 12.5 gives us

$$\begin{aligned}
\mathbb{E}_\tau \underline{\gamma}_y(\tau+1) &= \underline{\gamma}_y(\tau) + d_{\tau-y+1}\mathbb{E}_\tau[\gamma_y^{\tau-y+1}] \\
&= \underline{\gamma}_y(\tau) + d_{\tau-y+1}\rho M(y-1,\tau) \\
&= M(y,\tau) - (1-D_{\tau-y})\rho M(y-1,\tau) + d_{\tau-y+1}\mathbb{E}_\tau \rho M(y-1,\tau) \\
&= M(y,\tau) - (1-D_{\tau-y+1})\rho M(y-1,\tau)
\end{aligned}$$

where we have used Equation 12.31 to remove the dependence on $\underline{\gamma}_y(\tau)$.

It therefore follows that

$$\begin{aligned}
\mathbb{E}_\tau M(y,\tau+1) &= \sum_{s=0}^{\infty} \text{Þ}_{\tau+1-y,s}\rho^s \left(M(y-s,\tau) - (1-D_{\tau-y+1})\rho M(y-s-1,\tau)\right) \\
&= \sum_{s=0}^{\infty} \text{Þ}_{\tau+1-y,s}\rho^s M(y-s,\tau) - \sum_{s=0}^{\infty}(1-D_{\tau-y+1})\text{Þ}_{\tau+1-y,s}\rho^{s+1} M(y-s-1,\tau) \\
&= \sum_{s=0}^{\infty} \text{Þ}_{\tau+1-y,s}\rho^s M(y-s,\tau) - \sum_{s=0}^{\infty} \text{Þ}_{\tau+1-y,s+1}\rho^{s+1} M(y-s-1,\tau) \\
&= \text{Þ}_{\tau+1-y,0}\rho^0 M(y,\tau) \\
&= M(y,\tau)
\end{aligned}$$

as required.

Perhaps unsurprisingly, demonstrating Equation 12.30 is trickier. We start by showing that it is true when $y = \tau + 1 - X$, i.e., the cohort is one year away from being fully run off. Trivially $V(\tau+1-X,\tau+1) = 0$, since at time $\tau+1$, everyone in the cohort born at $\tau+1-X$ has died and so the cohort parameter $\gamma_{\tau+1-X} = \underline{\gamma}_{\tau+1-X}(\tau+1)$ is known

with certainty. Therefore

$$\begin{aligned}\mathbb{V}ar_\tau(M(\tau+1-X,\tau+1)) &= \mathbb{V}ar_\tau(\underline{\gamma}_{\tau+1-X}(\tau+1)) \\ &= \mathbb{V}ar_\tau(\underline{\gamma}_{\tau+1-x}(\tau)+d_X\gamma^X_{\tau+1-X}) \\ &= d_X^2\frac{\sigma^2}{d_X} \\ &= d_X\sigma^2 = (1-D_{X-1})\sigma^2 = V(\tau+1-X,\tau)\end{aligned}$$

using Equations 12.7 and 11.21. This is the first step in an induction argument, enabling us to work forwards in y to prove that Equation 12.30 holds true

$$\begin{aligned}\mathbb{V}ar_\tau(M(y,\tau+1)) &= \mathbb{V}ar_\tau\left(\underline{\gamma}_y(\tau+1)+(1-D_{\tau-y+1})\rho M(y-1,\tau+1)\right) \\ &= \mathbb{V}ar_\tau(\underline{\gamma}_y(\tau+1))+(1-D_{\tau-y+1})^2\rho^2\mathbb{V}ar_\tau(M(y-1,\tau+1)) \\ &\quad +2(1-D_{\tau-y+1})\rho\mathbb{C}ov_\tau(\underline{\gamma}_y(\tau+1),M(y-1,\tau+1))\end{aligned}$$

using Equation 12.31 and expanding the variance. Looking at the first of these parts, we see

$$\begin{aligned}\mathbb{V}ar_\tau(\underline{\gamma}_y(\tau+1)) &= \mathbb{V}ar_\tau(\underline{\gamma}_y(\tau)+d_{\tau-y+1}\gamma_y^{\tau-y+1}) \\ &= d^2_{\tau-y+1}\mathbb{V}ar_\tau(\gamma_y^{\tau-y+1}) \\ &= d_{\tau-y+1}\sigma^2+\rho^2d^2_{\tau-y+1}V(y-1,\tau)\end{aligned}$$

from Equation 12.7. For the second part, we assume that Equation 12.30 holds for $y-1$, using the inductive argument, and therefore

$$\mathbb{V}ar_\tau(M(y-1,\tau+1)) = V(y-1,\tau)-V(y-2,\tau)$$

Consequently

$$\begin{aligned}&\mathbb{V}ar_\tau(\underline{\gamma}_y(\tau+1))+(1-D_{\tau-y+1})^2\rho^2\mathbb{V}ar_\tau(M(y-1,\tau+1)) \\ &\quad= d_{\tau-y+1}\sigma^2+\rho^2\left(d^2_{\tau-y+1}+(1-D_{\tau-y+1})^2\right)V(y-1,\tau) \\ &\qquad-(1-D_{\tau-y+1})^2\rho^2V(y-2,\tau) \\ &\quad= d_{\tau-y+1}\sigma^2+\rho^2\left((1-D_{\tau-y+1}+d_{\tau-y+1})^2-2(1-D_{\tau-y+1})d_{\tau-y+1}\right)V(y-1,\tau) \\ &\qquad-(1-D_{\tau-y+1})\sigma^2-V(y-1,\tau)\quad\text{using Equation 12.32 on } V(y-2,\tau) \\ &\quad= (1-D_{\tau-y})\sigma^2+\rho^2(1-D_{\tau-y})^2V(y-1,\tau)-V(y-1,\tau) \\ &\qquad-2\rho^2(1-D_{\tau-y+1})d_{\tau-y+1}V(y-1,\tau) \\ &\quad= V(y,\tau)-V(y-1,\tau)-2\rho^2(1-D_{\tau-y+1})d_{\tau-y+1}V(y-1,\tau)\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{V}ar_\tau(M(y,\tau+1)) &= V(y,\tau) - V(y-1,\tau)\\ &+2(1-D_{\tau-y+1})\rho\left(\mathbb{C}ov_\tau(\underline{\gamma}_y(\tau+1), M(y-1,\tau+1)) - \rho d_{\tau-y+1}V(y-1,\tau)\right)\end{aligned}$$

and so Equation 12.30 will hold if and only if

$$\mathbb{C}ov_\tau(\underline{\gamma}_y(\tau+1), M(y-1,\tau+1)) = \rho d_{\tau-y+1}V(y-1,\tau)$$

To show that this calculation holds, we decompose the covariance as

$$\begin{aligned}\mathbb{C}ov_\tau(\underline{\gamma}_y(\tau+1), M(y-1,&\tau+1)) = d_{\tau+1-y}\mathbb{C}ov_\tau(\gamma_y^{\tau+1-y}, M(y-1,\tau+1))\\ &= d_{\tau+1-y}\sum_{s=0}^{\infty}\text{Þ}_{\tau-y+2,s}\rho^s\mathbb{C}ov_\tau(\gamma_y^{\tau+1-y}, \underline{\gamma}_{y-1-s}(\tau+1))\\ &= d_{\tau+1-y}\sum_{s=0}^{\infty}\text{Þ}_{\tau-y+2,s}\rho^s d_{\tau+2-y+2}\mathbb{C}ov_\tau(\gamma_y^{\tau+1-y}, \gamma_{y-s-1}^{\tau+2-y+s})\\ &= d_{\tau+1-y}\sum_{s=0}^{\infty}\text{Þ}_{\tau-y+2,s}\rho^s d_{\tau+2-y+s}\rho^{s+1}\text{Þ}_{\tau+1-y,s+1}\frac{\sigma^2}{d_{\tau+2-y+s}}\\ &\quad \text{from Equation 12.8}\\ &= \rho d_{\tau+1-y}\sum_{s=0}^{\infty}(1-D_{\tau+1-y+s})\text{Þ}^2_{\tau+1-y,s+1}\rho^{2s}\sigma^2\\ &= \rho d_{\tau+1-y}V(y-1,\tau)\end{aligned}$$

from the definition of $V(y,\tau)$ in Equation 11.21. Therefore, Equation 12.30 does indeed hold and models involving a set of cohort parameters are self-consistent in the real-world $\mathbb{P}$-measure.

12.A.2 Self-consistency in the market-consistent measure

Together, the results of Section 12.2.1 and Appendix 12.A.1 show that the forward mortality rates are self-consistent in the real-world $\mathbb{P}$-measure, as expected. We now demonstrate that they are self-consistent in the market-consistent $\mathbb{Q}$-measure, i.e.,

$$\mathbb{E}^{\mathbb{Q}}{}_\tau\, \nu^{\mathbb{Q}}_{x,t}(\tau+1) = \nu^{\mathbb{Q}}_{x,t}(\tau)$$

From Equation 11.31, we have

$$\begin{aligned}\nu^{\mathbb{Q}}_{x,t}(\tau+1) &= \exp\left(\boldsymbol{\beta}_x^\top \mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x})\right) \times \nu^{\mathbb{P}}_{x,t}(\tau+1) \\ &= \exp\Bigg(\alpha_x + \boldsymbol{\beta}_x^\top \mathbb{E}^{\mathbb{P}}{}_{\tau+1}\boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x^\top \mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{E}^{\mathbb{P}}{}_{\tau+1}\gamma_{t-x} \\ &\quad + \frac{1}{2}\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x}) + \boldsymbol{\beta}_x^\top \mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x})\Bigg)\end{aligned}$$

and also from Equation 11.29

$$\mathbb{E}^{\mathbb{Q}}{}_\tau\, \nu^{\mathbb{Q}}_{x,t}(\tau+1) = \frac{\mathbb{E}^{\mathbb{P}}{}_\tau\left[\exp\left(-\boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t - \lambda^\gamma\gamma_{t-x}\right)\nu^{\mathbb{Q}}_{x,t}(\tau+1)\right]}{\mathbb{E}^{\mathbb{P}}\exp\left(-\boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t - \lambda^\gamma\gamma_{t-x}\right)}$$

Looking first at the denominator

$$\left[\mathbb{E}^{\mathbb{P}}\exp\left(-\boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t - \lambda^\gamma\gamma_{t-x}\right)\right]^{-1} = \exp\left(\boldsymbol{\lambda}^\top\mathbb{E}^{\mathbb{P}}{}_\tau\boldsymbol{\kappa}_t - \frac{1}{2}\boldsymbol{\lambda}^\top\mathbb{V}ar^{\mathbb{P}}_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma\mathbb{E}^{\mathbb{P}}{}_\tau\gamma_{t-x} - \frac{1}{2}\lambda^{\gamma\,2}\mathbb{V}ar^{\mathbb{P}}_\tau(\gamma_{t-x})\right)$$

Next, let us consider the numerator

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}{}_\tau\left[\exp\left(-\boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t - \lambda^\gamma\gamma_{t-x}\right)\nu^{\mathbb{Q}}_{x,t}(\tau+1)\right] = \\ &\quad \exp\Bigg(\alpha_x + \boldsymbol{\beta}_x^\top\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x^\top\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{E}^{\mathbb{P}}{}_{\tau+1}\gamma_{t-x} \\ &\quad + \frac{1}{2}\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x}) + \boldsymbol{\beta}_x^\top\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x}) - \boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t - \lambda^\gamma\gamma_{t-x}\Bigg) \\ &= \exp\left(\alpha_x + \frac{1}{2}\boldsymbol{\beta}_x^\top\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \frac{1}{2}\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x}) + \boldsymbol{\beta}_x^\top\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma\mathbb{V}ar^{\mathbb{P}}_{\tau+1}(\gamma_{t-x})\right) \\ &\quad \times \mathbb{E}^{\mathbb{P}}{}_\tau\exp\left(\boldsymbol{\beta}_x\mathbb{E}^{\mathbb{P}}{}_{\tau+1}\boldsymbol{\kappa}_t - \boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t + \mathbb{E}^{\mathbb{P}}{}_{\tau+1}\gamma_{t-x} - \lambda^\gamma\gamma_{t-x}\right)\end{aligned}$$

Since all expectations and variances are under the measure $\mathbb{P}$ (unless stated otherwise), we drop the superscripts for simplicity. Considering only the expectation

$$\begin{aligned}&\mathbb{E}_\tau\exp\left(\boldsymbol{\beta}_x\mathbb{E}_{\tau+1}\boldsymbol{\kappa}_t - \boldsymbol{\lambda}^\top\boldsymbol{\kappa}_t + \mathbb{E}_{\tau+1}\gamma_{t-x} - \lambda^\gamma\gamma_{t-x}\right) = \\ &\quad \exp\Bigg(\boldsymbol{\beta}_x\mathbb{E}_\tau\boldsymbol{\kappa}_t - \boldsymbol{\lambda}^\top\mathbb{E}_\tau\boldsymbol{\kappa}_t + \mathbb{E}_\tau\gamma_{t-x} - \lambda^\gamma\mathbb{E}_\tau\gamma_{t-x} + \frac{1}{2}\boldsymbol{\beta}_x^\top\mathbb{V}ar_\tau(\mathbb{E}_{\tau+1}\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \frac{1}{2}\boldsymbol{\lambda}^\top\mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} \\ &\quad + \boldsymbol{\beta}_x^\top\mathbb{C}ov_\tau(\mathbb{E}_{\tau+1}\boldsymbol{\kappa}_t, \boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \frac{1}{2}\mathbb{V}ar_\tau(\mathbb{E}_{\tau+1}\gamma_{t-x}) + \frac{1}{2}\lambda^{\gamma\,2}\mathbb{V}ar_\tau(\gamma_{t-x}) - \lambda^\gamma\mathbb{C}ov_\tau(\mathbb{E}_{\tau+1}\gamma_{t-x}, \gamma_{t-x})\Bigg)\end{aligned}$$

Looking at each of the variance terms, we use the results

$$\begin{aligned}
\mathbb{V}ar_\tau(\mathbb{E}_{\tau+1}X) &= \mathbb{V}ar_\tau(X) - \mathbb{V}ar_{\tau+1}(X) \\
\mathbb{C}ov_\tau(X, \mathbb{E}_{\tau+1}X) &= \mathbb{E}_\tau \mathbb{C}ov_{\tau+1}(X, \mathbb{E}_{\tau+1}X) + \mathbb{C}ov_\tau(\mathbb{E}_{\tau+1}X, \mathbb{E}_{\tau+1}X) \\
&= 0 + \mathbb{V}ar_\tau(\mathbb{E}_{\tau+1}X) \\
&= \mathbb{V}ar_\tau(X) - \mathbb{V}ar_{\tau+1}(X)
\end{aligned}$$

to give

$$\begin{aligned}
&\mathbb{E}_\tau \exp\left(\boldsymbol{\beta}_x \mathbb{E}_{\tau+1}\boldsymbol{\kappa}_t - \boldsymbol{\lambda}^\top \boldsymbol{\kappa}_t + \mathbb{E}_{\tau+1}\gamma_{t-x} - \lambda^\gamma \gamma_{t-x}\right) = \\
&\quad \exp\left(\boldsymbol{\beta}_x \mathbb{E}_\tau \boldsymbol{\kappa}_t - \boldsymbol{\lambda}^\top \mathbb{E}_\tau \boldsymbol{\kappa}_t + \mathbb{E}_\tau \gamma_{t-x} - \lambda^\gamma \mathbb{E}_\tau \gamma_{t-x} + \frac{1}{2}\boldsymbol{\beta}_x^\top \left[\mathbb{V}ar_\tau(\boldsymbol{\kappa}_t) - \mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t)\right]\boldsymbol{\beta}_x\right. \\
&\quad + \frac{1}{2}\boldsymbol{\lambda}^\top \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \boldsymbol{\beta}_x^\top \left[\mathbb{V}ar_\tau(\boldsymbol{\kappa}_t) - \mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t)\right]\boldsymbol{\lambda} + \frac{1}{2}\mathbb{V}ar_\tau(\gamma_{t-x}) - \frac{1}{2}\mathbb{V}ar_{\tau+1}(\gamma_{t-x}) \\
&\quad \left. + \frac{1}{2}\lambda^{\gamma\,2}\mathbb{V}ar_\tau(\gamma_{t-x}) - \lambda^\gamma \mathbb{V}ar_\tau(\gamma_{t-x}) + \lambda^\gamma \mathbb{V}ar_{\tau+1}(\gamma_{t-x})\right)
\end{aligned}$$

Putting all three parts together and cancelling terms, we find

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}{}_\tau\, \nu^{\mathbb{Q}}_{x,t}(\tau+1) &= \exp\left(\alpha_x + \boldsymbol{\beta}_x^\top \mathbb{E}_\tau \boldsymbol{\kappa}_t + \frac{1}{2}\boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \mathbb{E}_\tau \gamma_{t-x} + \frac{1}{2}\mathbb{V}ar_\tau(\gamma_{t-x})\right. \\
&\qquad \left. + \boldsymbol{\beta}_x^\top \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar_\tau(\gamma_{t-x})\right) \\
&= \exp\left(\boldsymbol{\beta}_x^\top \Lambda \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\boldsymbol{\beta}_x + \lambda^\gamma \mathbb{V}ar_\tau(\gamma_{t-x})\right) \nu^{\mathbb{P}}_{x,t}(\tau) \\
&= \nu^{\mathbb{Q}}_{x,t}(\tau)
\end{aligned}$$

i.e., that forward mortality rates are self-consistent martingales under the market-consistent $\mathbb{Q}$-measure. From this, we also see that

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}{}_\tau \nu^{\mathbb{Q}}_{x,t}(\tau+1) &= \mathbb{E}^{\mathbb{P}}{}_\tau \exp\left(\boldsymbol{\beta}^\top \mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar_{\tau+1}(\gamma_{t-x})\right) \nu^{\mathbb{P}}_{x,t}(\tau+1) \\
&= \exp\left(\boldsymbol{\beta}^\top \mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t)\boldsymbol{\lambda} + \lambda^\gamma \mathbb{V}ar_{\tau+1}(\gamma_{t-x})\right) \nu^{\mathbb{P}}_{x,t}(\tau) \\
&= \exp\left(\boldsymbol{\beta}^\top \left[\mathbb{V}ar_{\tau+1}(\boldsymbol{\kappa}_t) - \mathbb{V}ar_\tau(\boldsymbol{\kappa}_t)\right]\boldsymbol{\lambda} + \lambda^\gamma \left[\mathbb{V}ar_{\tau+1}(\gamma_{t-x}) - \mathbb{V}ar_\tau(\gamma_{t-x})\right]\right) \nu^{\mathbb{Q}}_{x,t}(\tau)
\end{aligned}$$

i.e., the change of measure introduces a distortion which prevents market consistent forward rates being self-consistent in the real-world $\mathbb{P}$-measure.

Bibliography

Alai, D. H., Ignatieva, K., Sherris, M., 2013. Modelling longevity risk: Generalizations of the Olivier-Smith Model. Tech. rep., University of New South Wales.

Alai, D. H., Sherris, M., 2012. Rethinking age-period-cohort mortality trend models. Scandinavian Actuarial Journal 18 (2), 452–466.

Andreev, K. F., Vaupel, J., 2006. Forecasts of cohort mortality after age 50. Tech. rep., Max Planck Institute for Demographic Research.

Antolin, P., 2007. Longevity risk and private pensions. OECD Working Papers.

Arnold (-Gaille), S., Sherris, M., 2013. Forecasting mortality trends allowing for cause-of-death mortality dependence. North American Actuarial Journal 17 (4), 273–282.

Aro, H., 2014. Systematic and nonsystematic mortality risk in pension portfolios. North American Actuarial Journal 18 (1), 59–67.

Aro, H., Pennanen, T., 2011. A user-friendly approach to stochastic mortality modelling. European Actuarial Journal 1 (S2), 151–167.

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66 (1), 47–78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18 (1), 1–22.

Ballotta, L., Haberman, S., 2006. The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case. Insurance: Mathematics and Economics 38 (1), 195–214.

Barbarin, J., 2008. Heath-Jarrow-Morton modelling of longevity bonds and the risk minimization of life insurance portfolios. Insurance: Mathematics and Economics 43 (1), 41–55.

Barrieu, P. M., Bensusan, H., Karoui, N. E., Hillairet, C., Loisel, S., Ravanelli, C., 2012. Understanding, modelling and managing longevity risk: Key issues and main challenges. Scandinavian Actuarial Journal 3, 203–231.

Bauer, D., Bergmann, D., Reuß, A., 2009. Solvency II and nested simulations - A least-squares Monte Carlo approach. Tech. rep., University of Ulm.

Bauer, D., Börger, M., Ruß, J., Zwiesler, H., 2008. The volatility of mortality. Asia-Pacific Journal of Risk and Insurance 3 (10), 2153–3792.

Bauer, D., Kramer, F., 2007. Risk and valuation of mortality contingent catastrophe bonds. Tech. rep., University of Ulm.

Bauer, D., Reuß, A., Singer, D., 2012. On the calculation of the solvency capital requirement based on nested simulations. ASTIN Bulletin 42 (2), 453–499.

Bayraktar, E., Young, V. R., 2007. Hedging life insurance with pure endowments. Insurance: Mathematics and Economics 40 (3), 435–444.

Beelders, O., Colarossi, D., 2004. Modelling mortality risk with extreme value theory: The case of Swiss Re's mortality-indexed bonds. Global Association of Risk Professionals, 26–30.

Biffis, E., Blake, D., Pitotti, L., Sun, A., 2014. The cost of counterparty risk and collateralization in longevity swaps. Journal of Risk and Insurance, Forthcoming.

Blake, D., Burrows, W., 2001. Survivor bonds: Helping to hedge mortality risk. Journal of Risk and Insurance 68 (2), 339–348.

Blake, D., Cairns, A. J. G., Coughlan, G. D., Dowd, K., MacMinn, R., 2013. The new life market. Journal of Risk and Insurance 80 (3), 501–558.

Blake, D., Cairns, A. J. G., Dowd, K., 2006. Living with mortality: Longevity bonds and other mortality-linked securities. British Actuarial Journal 12 (1), 153–197.

Blake, D., Harrison, D., 2013. A healthier way to de-risk: The introduction of medical underwriting to the defined benefit de-risking market. Tech. rep., Pensions Institute, Cass Business School, City University London.

Booth, H., 2006. Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting 22 (3), 547–581.

Booth, H., Maindonald, J., Smith, L., 2002. Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56 (3), 325–36.

Börger, M., 2010. Deterministic shock vs. stochastic value-at-risk - An analysis of the Solvency II standard model approach to longevity risk. Blätter der DGVFM 31 (2), 225–259.

Börger, M., Fleischer, D., Kuksin, N., 2013. Modeling the mortality trend under modern solvency regimes. ASTIN Bulletin 44 (1), 1–38.

Börger, M., Ruß, J., 2012. It takes two: Why mortality trend modeling is more than modeling one mortality trend. Tech. rep., University of Ulm.

Brouhns, N., Denuit, M. M., Van Keilegom, I., 2005. Bootstrapping the Poisson log-bilinear model for mortality forecasting. Scandinavian Actuarial Journal 2005 (3), 212–224.

Brouhns, N., Denuit, M. M., Vermunt, J., 2002a. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31 (3), 373–393.

Brouhns, N., Denuit, M. M., Vermunt, J., 2002b. Measuring the longevity risk in mortality projections. Bulletin of the Swiss Association of Actuaries 2, 105–130.

Cairns, A. J. G., 2000. A discussion of parameter and model uncertainty in insurance. Insurance: Mathematics and Economics 27 (3), 313–330.

Cairns, A. J. G., 2007. A multifactor generalisation of the Olivier-Smith model for stochastic mortality. Tech. rep., Heriot-Watt University, Edinburgh.

Cairns, A. J. G., 2011. Modelling and management of longevity risk: Approximations to survivor functions and dynamic hedging. Insurance: Mathematics and Economics 49 (3), 438–453.

Cairns, A. J. G., 2013. Robust hedging of longevity risk. Journal of Risk and Insurance 80 (3), 621–648.

Cairns, A. J. G., Blake, D., Dowd, K., 2006a. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. Journal of Risk and Insurance 73 (4), 687–718.

Cairns, A. J. G., Blake, D., Dowd, K., 2006b. Pricing death: Frameworks for the valuation and securitization of mortality risk. ASTIN Bulletin 36 (1), 79–120.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2011a. 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.

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Khalaf-Allah, M., 2011b. Bayesian stochastic mortality modelling for two populations. ASTIN Bulletin 41 (1), 29–59.

Cairns, A. J. G., Blake, D., Dowd, K., Kessler, A., 2014. Phantoms never die: Living with unreliable mortality data. Tech. rep., Herriot Watt University, Edinburgh.

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

Callot, L., Haldrup, N., Lamb, M. K., 2014. Deterministic and stochastic trends in the Lee-Carter mortality model. Tech. rep., Aarhus University.

Campos, J., Ericsson, N. R., Hendry, D. F., 2005. General-to-specific modeling: An overview and selected bibliography. Federal Reserve Discussion Papers.

Carstensen, B., 2007. Age-period-cohort models for the Lexis diagram. Statistics in Medicine 26, 3018–3045.

Carter, L. R., Prskawetz, A., 2001. Examining structural shifts in mortality using the Lee-Carter method. Tech. rep., Max Planck Institute for Demographic Research.

Carter, R., Lee, D., 1992. Modeling and forecasting US sex differentials. International Journal of Forecasting 8, 393–411.

Chen, H., Cox, S. H., 2009. Modeling mortality with jumps: Applications to mortality securitization. Journal of Risk and Insurance 76 (3), 727–751.

Christiansen, M., Niemeyer, A., 2014. Fundamental definition of the Solvency Capital Requirement in Solvency II. ASTIN Bulletin 44 (03), 501–533.

Clayton, D., Schifflers, E., 1987. Models for temporal variation in cancer rates. II: Age-period-cohort models. Statistics in Medicine 6 (4), 469–81.

Coelho, E., Nunes, L. C., 2011. Forecasting mortality in the event of a structural change. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (3), 713–736.

Continuous Mortality Investigation, 2002. Working Paper 1 - An interim basis for adjusting the "92" series mortality projections for cohort effects.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-1

Continuous Mortality Investigation, 2004. Working Paper 3 - Projecting future mortality: A discussion paper.
URL http://www.actuaries.org.uk/research-and-resources/documents/cmi-working-paper-3-pr

Continuous Mortality Investigation, 2007. Working Paper 25 - Stochastic projection methodologies: Lee-Carter model features, example results and implications.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-25

Continuous Mortality Investigation, 2008. Working Paper 35 - The graduations of the CMI self-administered pension schemes 2000-2006 mortality experience: Final "S1" series of mortality tables.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-papers-34-and-

Continuous Mortality Investigation, 2009a. Working Paper 38 - A prototype mortality projection model: Part one - an outline of the proposed approach.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-papers-38-and-

Continuous Mortality Investigation, 2009b. Working Paper 39 - A prototype mortality projections model: Part two - detailed analysis.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-papers-38-and-

Continuous Mortality Investigation, 2011. Working Paper 53 - An initial investigation into rates of mortality improvement for pensioners of self-administered pension schemes.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-53

Continuous Mortality Investigation, 2012. Working Paper 61 - An investigation into the mortality experience by industry classification of pensioners.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-61

Continuous Mortality Investigation, 2013. Working Paper 69 - The CMI mortality projections model, CMI_2013, and feedback on the consultation on the future of the CMI Library of Mortality Projections and the CMI Mortality Projections Model.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-69

Continuous Mortality Investigation, 2014a. Working Paper 71 - Graduations of the CMI SAPS 2004-2011 mortality experience based on data collected by 30 June 2012: Final "S2" series of mortality tables.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-71

Continuous Mortality Investigation, 2014b. Working Paper 73 - Analysis of the mortality experience of pensioners of self-administered pension schemes for the period 2005 to 2012.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-73

Continuous Mortality Investigation, 2014c. Working Paper 76 - Analysis of the mortality experience of pensioners of self-administered pension schemes for the period 2006 to 2013.
URL http://www.actuaries.org.uk/research-and-resources/pages/cmi-working-paper-76

Cossette, H., Delwarde, A., Denuit, M. M., Guillot, F., Marceau, E., 2007. Pension plan valuation and mortality projection: A case study with mortality data. North American Actuarial Journal 11 (2), 1–34.

Coughlan, G. D., Epstein, D., Ong, A., Sinha, A., 2007a. LifeMetrics: A toolkit for measuring and managing longevity and mortality risks. Technical document. JPMorgan Pension Advisory Group.

Coughlan, G. D., Epstein, D., Sinha, A., Honig, P., 2007b. q-forwards: Derivatives for transferring longevity and mortality risks. JPMorgan Pension Advisory Group.

Coughlan, G. D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A. J. G., Blake, D., Dowd, K., 2011. Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. North American Actuarial Journal 15 (2), 150–176.

Cowley, A., Cummins, J., 2005. Securitization of life insurance assets and liabilities. Journal of Risk and Insurance 72 (2), 193–226.

Cox, D. R., 1972. Regression models and life tables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 34, 187–202.

Cox, S. H., Lin, Y., 2007. Natural hedging of life and annuity mortality risks. North American Actuarial Journal 11 (3), 1–15.

Cox, S. H., Lin, Y., Wang, S., 2006. Multivariate exponential tilting and pricing implications for mortality securitization. Journal of Risk and Insurance 73 (4), 719–736.

Currie, I. D., 2014. On fitting generalized linear and non-linear models of mortality. Scandinavian Actuarial Journal, Forthcoming.

Currie, I. D., Durbán, M., Eilers, P., 2004. Smoothing and forecasting mortality rates. Statistical Modelling 4 (4), 279–298.

Czado, C., Delwarde, A., Denuit, M. M., 2005. Bayesian Poisson log-bilinear mortality projections. Insurance: Mathematics and Economics 36 (3), 260–284.

D'Amato, V., Lorenzo, E. D., Haberman, S., Russolillo, M., Sibillo, M., 2011. The Poisson log-bilinear Lee-Carter model: Applications of efficient bootstrap methods to annuity analyses. North American Actuarial Journal 15 (2), 315–333.

Darkiewicz, G., Hoedemakers, T., 2004. How the co-integration analysis can help in mortality forecasting. Tech. rep., Catholic University of Leuven.

Dawson, P., Dowd, K., Cairns, A. J. G., Blake, D., 2010. Survivor derivatives: A consistent pricing framework. Journal of Risk and Insurance 77 (3), 579–596.

de Grey, A. D. N. J., 2006. Extrapolaholics anonymous: Why demographers' rejections of a huge rise in cohort life expectancy in this century are overconfident. Annals of the New York Academy of Sciences 1067, 83–93.

Debón, A., Martinez-Ruiz, F., Montes, F., Martínez-Ruiz, F., Montes, F., 2010. A geostatistical approach for dynamic life tables: The effect of mortality on remaining lifetime and annuities. Insurance: Mathematics and Economics 47 (3), 327–336.

Debón, A., Montes, F., Mateu, J., Porcu, E., Bevilacqua, M., 2008. Modelling residuals dependence in dynamic life tables: A geostatistical approach. Computational Statistics & Data Analysis 52 (6), 3128–3147.

Delwarde, A., Denuit, M. M., Eilers, P., 2007a. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. Statistical Modelling 7 (1), 29–48.

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

Delwarde, A., Denuit, M. M., Partrat, C., 2007b. Negative binomial version of the Lee-Carter model for mortality forecasting. Applied Stochastic Models in Business and Industry 23 (5), 385–401.

Denuit, M. M., 2008. Comonotonic approximations to quantiles of life annuity conditional expected present values. Insurance: Mathematics and Economics 42 (1), 831–838.

Denuit, M. M., 2009. An index for longevity risk transfer. Journal of Computational and Applied Mathematics 230 (2), 411–417.

Denuit, M. M., Devolder, P., Goderniaux, A.-M., 2007. Securitization of longevity risk: Pricing survivor bonds with Wang transform in the Lee-Carter framework. Journal of Risk and Insurance 74 (1), 87–113.

Denuit, M. M., Dhaene, J., Goovaerts, M., Kaas, R., 2005. Actuarial theory for dependent risks: Measures, orders and models. Wiley.

Denuit, M. M., Goderniaux, A.-M., 2005. Closing and projecting life tables using log-linear models. Bulletin of the Swiss Association of Actuaries 1, 29–49.

Dhaene, J., Kukush, A., Luciano, E., Schoutens, W., Stassen, B., 2013. On the (in-)dependence between financial and actuarial risks. Insurance: Mathematics and Economics 52 (3), 522–531.

Donnelly, C., 2014. Quantifying mortality risk in small defined-benefit pension schemes. Scandinavian Actuarial Journal (1), 41–57.

Dowd, K., 2003. Survivor bonds: A comment on Blake and Burrows. Journal of Risk and Insurance 70 (2), 339–348.

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

Dowd, K., Blake, D., Cairns, A. J. G., 2011a. A computationally efficient algorithm for estimating the distribution of future annuity values under interest-rate and longevity risks. North American Actuarial Journal 15 (2), 237–247.

Dowd, K., Blake, D., Cairns, A. J. G., Dawson, P., 2006a. Survivor swaps. Journal of Risk and Insurance 73 (1), 1–17.

Dowd, K., Cairns, A. J. G., Blake, D., 2006b. Mortality-dependent financial risk measures. Insurance: Mathematics and Economics 38 (3), 427–440.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., 2011b. A gravity model of mortality rates for two related populations. North American Actuarial Journal 15 (2), 334–356.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2010b. Backtesting stochastic mortality models: An ex post evaluation of multiperiod-ahead density forecasts. North American Actuarial Journal 13 (3), 281–298.

Dowd, K., Cairns, A. J. G., Blake, D., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2010c. Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics 47 (3), 255–265.

Durbin, J., Watson, G. S., 1951. Testing for serial correlations in least squares regression. Biometrika 38 (1/2), 159–177.

EIOPA, 2014. Technical specification for the preparatory phase (Part I). Tech. rep., European Insurance and Occupational Pensions Authority, Frankfurt.

Fetiveau, C., Jia, C., 2014. Longevity risk hedging with population based index solution - A study of basis risk based on England & Wales population. Tech. rep., Deutsche Bank.

Fienberg, S. E., Mason, W. M., 1979. Identification and estimation of age-period-cohort models in the analysis of discrete archival data. Sociological Methodology 10, 1–67.

Finkelstein, M., 2012. Discussing the Strehler-Mildvan model of mortality. Demographic Research 26, 191–206.

Frederic, P., Lad, F., 2008. Two moments of the logitnormal distribution. Communications in Statistics - Simulation and Computation 37 (7), 1263–1269.

French, D., 2014. International mortality modelling - An economic perspective. Economics Letters 122 (2), 182–186.

Gaille, S., Sherris, M., 2011. Modelling mortality with common stochastic long-run trends. The Geneva Papers on Risk and Insurance Issues and Practice 36 (4), 595–621.

Gavrilov, L. A., Gavrilova, N. S., 2011. Mortality measurement at advanced ages: A study of the social security administration death master file. North American Actuarial Journal 15 (3), 432–447.

Gerber, H., Shiu, E., 1994. Option pricing by Esscher transforms. Transactions of the Society of Actuaries 46, 99–191.

Glenn, N., 1976. Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. American Sociological Review 41 (5), 900–904.

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Gutterman, S., Vanderhoof, I. T., 1998. Forecasting changes in mortality. North American Actuarial Journal 2 (4), 135–138.

Haberman, S., Kaishev, V. K., Millossovich, P., Villegas, A. M., Baxter, S. D., Gaches, A. T., Gunnlaugsson, S., Sison, M., 2014. Longevity basis risk: A methodology for assessing basis risk. Tech. rep., Cass Business School, City University London and Hymans Robertson LLP.

Haberman, S., Renshaw, A., 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.

Haberman, S., Renshaw, A., 2012. Parametric mortality improvement rate modelling and projecting. Insurance: Mathematics and Economics 50 (3), 309–333.

Haberman, S., Renshaw, A., 2013. Modelling and projecting mortality improvement rates using a cohort perspective. Insurance: Mathematics and Economics 53 (1), 150–168.

Hainaut, D., 2012. Multidimensional Lee-Carter model with switching mortality processes. Insurance: Mathematics and Economics 50 (2), 236–246.

Hanewald, K., 2011. Explaining mortality dynamics: The role of macroeconomic fluctuations and cause of deaths trends. North American Actuarial Journal 15 (2), 290–314.

Harper, S., Lynch, J., Burris, S., Davey Smith, G., 2007. Trends in the black-white life expectancy gap in the United States, 1983-2003. The Journal of the American Medical Association 297 (11), 1224–32.

Harris, D., Harvey, D. I., Leybourne, S. J., Taylor, A. R., 2009. Testing for a unit root in the presence of a possible break in trend. Econometric Theory 25 (06), 1545.

Hatzopoulos, P., Haberman, S., 2009. A parameterized approach to modeling and forecasting mortality. Insurance: Mathematics and Economics 44 (1), 103–123.

Hatzopoulos, P., Haberman, S., 2011. A dynamic parameterization modeling for the age-period-cohort mortality. Insurance: Mathematics and Economics 49 (2), 155–174.

Heligman, L., Pollard, J., 1980. The age pattern of mortality. Journal of the Institute of Actuaries 107 (1), 49–80.

Hendry, D. F., Massmann, M., 2005. Co-breaking: Recent advances and a synopsis of the literature. Tech. Rep. 1978, University of Oxford.

Hobcraft, J., Menken, J., Preston, S. H., 1982. Age, period and cohort effects in demography: A review. Population Index 48 (1), 4–43.

Holford, T. R., 1983. The estimation of age, period and cohort effects for vital rates. Biometrics 39 (2), 311–24.

Holland, B., Ahsanullah, M., 1989. Further results on a distribution of Meinhold and Singpurwalla. American Statistical Association 43 (4), 216–219.

Huang, J. Z., Shen, H., Buja, A., 2009. The analysis of two-way functional data using two-way regularized singular value decompositions. Journal of the American Statistical Association 104 (448), 1609–1620.

Human Mortality Database, 2014. Human Mortality Database. Tech. rep., University of California, Berkeley and Max Planck Institute for Demographic Research.
URL www.mortality.org

Hunt, A., Villegas, A. M., 2015. Robustness and convergence in the Lee-Carter model with cohort effects. Insurance: Mathematics and Economics 64, 186–202.

Hymans Robertson, 2015. Buy-outs, buy-ins and longevity hedging Q4 2014. Tech. rep., Hymans Robertson LLP.
URL http://www.hymans.co.uk/media/591924/150317-managing-pension-scheme-risk-q4-2014.pd

Hyndman, R., Ullah, M., 2007. Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.

Hyndman, R. J., Booth, H., Yasmeen, F., 2013. Coherent mortality forecasting: The product-ratio method with functional time series models. Demography 50, 261–283.

Institute for Fiscal Studies, 2014. Living standards, poverty and inequality in the UK: 2013. Tech. rep., Institute for Fiscal Studies.
URL http://www.ifs.org.uk/comms/r81.pdf

Jarner, S. F., Kryger, E. M., 2011. Modelling mortality in small populations: The SAINT model. ASTIN Bulletin 41 (2), 377–418.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59 (6), 1551–1580.

Juselius, K., 2006. The cointegrated VAR model: Methodology and applications. Oxford University Press.

Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. M., 2001. Modern actuarial risk theory. Kluwer Academic Publishers.

Kessler, A., 2014. Longevity risk and reinsurance: Strategies for managing annuity blocks. Tech. rep., Prudential Retirement.
URL http://www.cass.city.ac.uk/__data/assets/pdf_file/0018/232614/Longevity-10-Longevit

Keyfitz, N., 1985. Applied mathematical demography. Springer-Verlag.

Kijima, M., 2005. A multivariate extension of equilibrium pricing transforms: The multivariate Esscher and Wang transforms for pricing financial and insurance risks. Tech. rep., Kyoto University.

Kleinow, T., Cairns, A. J. G., 2013. Mortality and smoking prevalence: An empirical investigation in ten developed countries. British Actuarial Journal 18 (2), 452–466.

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.

Kuang, D., Nielsen, B., Nielsen, J. P., 2008a. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 987–991.

Kuang, D., Nielsen, B., Nielsen, J. P., 2008b. Identification of the age-period-cohort model and the extended chain-ladder model. Biometrika 95 (4), 979–986.

Lane, M., 2011. Longevity risk from the perspective of the ILS markets. The Geneva Papers on Risk and Insurance Issues and Practice 36 (4), 501–515.

Lane, M. N., Beckwith, R., 2011. Prague Spring or Louisiana morning? Annual review for the four quarters, Q2 2010 to Q1 2011. Tech. rep., Lane Finacial LLC.

Lane, M. N., Beckwith, R., 2012. More return; More risk: Annual review for the four quarters, Q2 2011 to Q1 2012. Tech. rep., Lane Financial LLC.

Lane, M. N., Beckwith, R., 2013. Soft markets ahead!? Annual review for the four quarters, Q2 2012 to Q1 2013. Tech. rep., Lane Financial LLC.

Lane, M. N., Beckwith, R., 2014. Straw hats in winter: Annual review for the four quarters, Q2 2013 to Q1 2014. Tech. rep., Lane Financial LLC.

Lazar, D., Denuit, M. M., 2009. A multivariate time series approach to projected life tables. Applied Stochastic Models in Business and Industry 25 (6), 806–823.

Lee, R. D., 2000. The Lee-Carter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal 4 (1), 80–93.

Lee, R. D., Carter, L. R., 1992. Modeling and forecasting U.S. mortality. Journal of the American Statistical Association 87 (419), 659– 671.

Lemoine, K., 2014. Mortality regimes and longevity risk in a life annuity portfolio. Scandinavian Actuarial Journal, Forthcoming.

Li, J., 2014. A quantitative comparison of simulation strategies for mortality projection. Annals of Actuarial Science 8 (02), 281–297.

Li, J. S.-H., Chan, W., 2005. Outlier analysis and mortality forecasting: The United Kingdom and Scandinavian countries. Scandinavian Actuarial Journal 2005 (3), 187–211.

Li, J. S.-H., Chan, W., Cheung, S., 2011. Structural changes in the Lee-Carter mortality indexes: Detection and implications. North American Actuarial Journal 15 (1), 13–31.

Li, J. S.-H., Hardy, M. R., 2011. Measuring basis risk in longevity hedges. North American Actuarial Journal 15 (2), 177–200.

Li, J. S.-H., Hardy, M. R., Tan, K. S., 2009. Uncertainty in mortality forecasting: An extension to the classical Lee-Carter approach. ASTIN Bulletin 39 (1), 137–164.

Li, J. S.-H., Luo, A., 2012. Key Q-duration: A framework for hedging longveity risk. ASTIN Bulletin 42 (2), 413–452.

Li, J. S.-h., Zhou, R., Hardy, M. R., 2015. A step-by-step guide to building two-population stochastic mortality models. Insurance: Mathematics and Economics 63, 121–134.

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

Li, N., Lee, R. D., Tuljapurkar, S., 2004. Using the Lee-Carter method to forecast mortality for populations with limited data. International Statistical Review 72 (1), 19–36.

Lin, Y., Cox, S. H., 2005. Securitization of mortality risks in life annuities. Journal of Risk and Insurance 72 (2), 227–252.

Lin, Y., Cox, S. H., 2008. Securitization of catastrophe mortality risks. Insurance: Mathematics and Economics 42 (2), 628–637.

Liu, X., Braun, W. J., 2010. Investigating mortality uncertainty using the block bootstrap. Journal of Probability and Statistics, Article ID 813583.

Liu, Y., Li, J. S.-H., 2015. The age pattern of transitory mortality jumps and its impact on the pricing of catastrophic mortality bonds. Insurance: Mathematics and Economics 64, 135–150.

Loeys, J., Panigirtzoglou, N., Ribeiro, R., 2007. Longevity: A market in the making. JPMorgan Pension Advisory Group.

Lu, J. L. C., Wong, W., Bajekal, M., 2012. Mortality improvement by socio-economic circumstances in England (1982 to 2006). British Actuarial Journal 19 (1), 1–35.

Manton, K. G., Patrick, C., Stallard, E., 1980. Mortality model based on delays in progression of chronic diseases: Alternative to cause elimination model. Public Health Reports 95, 580–588.

Mavros, G., Cairns, A. J. G., Kleinow, T., Streftaris, G., 2014. A parsimonious approach to stochastic mortality modelling with dependent residuals. Tech. rep., Heriot-Watt University, Edinburgh.

McCullagh, P., Nelder, J., 1983. Generalized linear models. Chapman and Hall.

Michaelson, A., Mulholland, J., 2014. Strategy for increasing the global capacity for longevity risk transfer: Developing transactions that attract capital markets investors. Journal of Alternative Investments 17 (1), 18–27.

Milevsky, M., Promislow, S., 2001. Mortality derivatives and the option to annuitise. Insurance: Mathematics and Economics 29, 299–318.

Milevsky, M., Promislow, S., Young, V. R., 2006. Killing the law of large numbers: Mortality risk premiums and the Sharpe ratio. Journal of Risk and Insurance 73 (4), 673–686.

Milevsky, M., Promislow, S., Young, V. R., Bayraktar, E., 2005. Financial valuation of mortality risk via the instantaneous Sharpe ratio. Journal of Economic Dynamics and Control 33 (3), 676–691.

Milidonis, A., Lin, Y., Cox, S. H., 2011. Mortality regimes and pricing. North American Actuarial Journal 15 (2), 266–289.

Miltersen, K. R., Persson, S.-A., 2005. Is mortality dead? Stochastic forward force of mortality rate determined by no arbitrage. Tech. rep., University of Ulm.

Mitchell, D., Brockett, P. L., Mendoza-Arriaga, R., Muthuraman, K., 2013. Modeling and forecasting mortality rates. Insurance: Mathematics and Economics 52 (2), 275–285.

Murphy, K. M., Topel, R. H., 2002. Estimation and inference in two-step econometirc models. Journal of Business & Economic Statistics 20 (1), 88–97.

Murphy, M., 2009. The "golden generations" in historical context. British Actuarial Journal 15 (S1), 151–184.

Murphy, M., 2010. Re-examining the dominance of birth cohort effects on mortality. Population and Development Review 36 (2), 365–90.

Nielsen, B., Nielsen, J. P., 2014. Identification and forecasting in mortality models. The Scientific World Journal, Article ID 347043.

Nielsen, L., 2010. Assessment of the VaR (99.5%) for longevity risk. Tech. rep., SamPension.

Norberg, R., 2010. Forward mortality and other vital rates - Are they the way forward? Insurance: Mathematics and Economics 47 (2), 105–112.

O'Brien, R. M., 2000. Age period cohort characteristic models. Social Science Research 29 (1), 123–139.

O'Brien, R. M., 2011. Constrained estimators and age-period-cohort models. Sociological Methods & Research 40 (3), 419–452.

Oeppen, J., Vaupel, J., 2002. Broken limits to life expectancy. Science 296 (5570), 1029–1031.

O'Hare, C., Li, Y., 2012a. Explaining young mortality. Insurance: Mathematics and Economics 50 (1), 12–25.

O'Hare, C., Li, Y., 2012b. Identifying structural breaks in stochastic mortality models. Tech. rep., Monash University, Melbourne.

Olivier, P., Jeffrey, T., 2004. Stochastic mortality models.
URL http://www.actuaries.ie/EventsandPapers/Events2004/2004-06-01_PensionerMortality/20

Olshansky, S., Carnes, B. A., Grahn, D., 1998. Confronting the boundaries of human longevity. American Scientist 86 (1), 52–61.

Olshansky, S., Douglas, J., Hershow, R., 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, 1138–1145.

Oppers, S. E., Chikada, K., Eich, F., Imam, P., Kiff, J., Kisser, M., Soto, M., Kim, Y. S., 2012. The financial impact of longevity risk. Tech. rep., International Monetary Fund.

Osmond, C., 1985. Using age, period and cohort models to estimate future mortality rates. International Journal of Epidemiology 14 (1), 124–9.

Pedroza, C., 2006. A Bayesian forecasting model: Predicting US male mortality. Biostatistics 7 (4), 530–550.

Pelsser, A., 2003. Pricing and hedging guaranteed annuity options via static option replication. Insurance: Mathematics and Economics 33 (2), 283–296.

Pitacco, E., Denuit, M. M., Haberman, S., Olivieri, A., 2009. Modelling longevity dynamics for pensions and annuity business. Oxford University Press.

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.

Plat, R., 2011. One-year value-at-risk for longevity and mortality. Insurance: Mathematics and Economics 49 (3), 462–470.

Reichmuth, W., Sarferaz, S., 2008. Bayesian demographic modeling and forecasting: An application to U.S. mortality. Tech. rep., Humbolt University, Berlin.

Renshaw, A., Haberman, S., 2003a. Lee-Carter mortality forecasting: A parallel generalized linear modelling approach for England and Wales mortality projections. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52 (1), 119–137.

Renshaw, A., Haberman, S., 2003b. Lee-Carter mortality forecasting with age-specific enhancement. Insurance: Mathematics and Economics 33 (2), 255–272.

Renshaw, A., Haberman, S., 2003c. On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics 32 (3), 379–401.

Renshaw, A., 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., 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.

Renshaw, A., Haberman, S., Hatzopoulos, P., 1996. The modelling of recent mortality trends in United Kingdom male assured lives. British Actuarial Journal 2 (2), 449–477.

Richards, S. J., 2008. Detecting year-of-birth mortality patterns with limited data. Journal of the Royal Statistical Society: Series A (Statistics in Society) 171 (1), 279–298.

Richards, S. J., Currie, I. D., Ritchie, G. P., 2014. A value-at-risk framework for longevity trend risk. British Actuarial Journal 19 (1), 116–139.

Richards, S. J., Kaufhold, K., Rosenbusch, S., 2013. Creating portfolio-specific mortality tables: A case study. European Actuarial Journal 3 (2), 295–319.

Riebler, A., Held, L., Rue, H. v., 2012. Estimation and extrapolation of time trends in registry data - Borrowing strength from related populations. The Annals of Applied Statistics 6 (1), 304–333.

Rodgers, W., 1982. Estimable functions of age, period, and cohort effects. American Sociological Review 47 (6), 774–787.

Ruhm, C., 2000. Are recessions good for your health? The Quarterly Journal of Economics 115 (2), 617–650.

Ruhm, C., 2004. Macroeconomic conditions, health and mortality. NBER Working Papers.

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., 2009. Longevity basis risk modeling: A co-integration based approach. Tech. rep., University of Lyon.

Schrager, D., 2006. Affine stochastic mortality. Insurance: Mathematics and Economics 38 (1), 81–97.

Shkolnikov, V., Andreev, E., Begun, A. Z., 2003. Gini coefficient as a life table function. Demographic Research 8, 305–358.

Siegel, J. S., 2005. The great debate on the outlook for human longevity: Exposition and evaluation of two divergent views. Tech. rep., Society of Actuaries.

Sithole, T., Haberman, S., Verrall, R., 2000. An investigation into parametric models for mortality projections, with applications to immediate annuitants' and life office pensioners' data. Insurance: Mathematics and Economics 27 (3), 285–312.

Sithole, T., Haberman, S., Verrall, R., 2012. Second international comparative study of mortality tables for pension fund retirees: A discussion paper. British Actuarial Journal 7 (3), 650–671.

Smith, A., 2005. Stochastic mortality modelling.
URL http://www.icms.org.uk/archive/meetings/2005/quantfinance/sci_prog.html

Standard and Poor's, 2010. Presale information: Kortis Capital Ltd. Tech. rep., Standard and Poors.

Stevens, R., De Waegenaere, A., Melenberg, B., 2010. Calculating capital requirements for longevity risk in life insurance products using an internal model in line with Solvency II. Tech. rep., University of Tilburg.

Sweeting, P. J., 2011. A trend-change extension of the Cairns-Blake-Dowd model. Annals of Actuarial Science 5 (02), 143–162.

Tan, C. I., Li, J., Li, J. S.-H., Balasooriya, U., 2014. Parametric mortality indexes: From index construction to hedging strategies. Insurance: Mathematics and Economics 59, 285–299.

Tappe, S., Weber, S., 2013. Stochastic mortality models: An infinite-dimensional approach. Finance and Stochastics 18 (1), 209–248.

The Pensions Regulator, 2013a. Scheme funding. Tech. rep.
URL http://www.thepensionsregulator.gov.uk/codes/code-funding-defined-benefits.aspx

The Pensions Regulator, 2013b. The Purple Book. Tech. rep.
URL http://www.thepensionsregulator.gov.uk/doc-library/research-analysis.aspx

Tuljapurkar, S., Edwards, R., 2009. Variance in death and its implications for modeling and forecasting mortality. NBER Working Papers.

Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405 (6788), 789–792.

van Berkum, F., Antonio, K., Vellekoop, M. H., 2014. The impact of multiple structural changes on mortality predictions. Scandinavian Actuarial Journal, Forthcoming.

Vaupel, J., Manton, K. G., Stallard, E., 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16 (3), 439–454.

Villegas, A. M., 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.

Wan, C., Bertschi, L., 2015. Swiss coherent mortality model as a basis for developing longevity de-risking solutions for Swiss pension funds: A practical approach. Insurance: Mathematics and Economics 63, 66–75.

Wang, C.-W., Huang, H., Liu, I.-C., 2011. A quantitative comparison of the Lee-Carter model under different types of non-Gaussian innovations. The Geneva Papers on Risk and Insurance Issues and Practice 36 (4), 675–696.

Wang, H., Preston, S. H., 2009. Forecasting United States mortality using cohort smoking histories. Proceedings of the National Academy of Sciences of the United States of America 106 (2), 393–8.

Wang, J. L., Huang, H., Yang, S. S., Tsai, J. T., 2009. An optimal product mix for hedging longevity risk in life insurance companies: The immunization theory approach. Journal of Risk and Insurance 77 (2), 473–497.

Wang, S., 2000. A class of distortion operators for pricing financial and insurance risks. Journal of Risk and Insurance 67 (1), 15–36.

Wang, S., 2002. A univeral framework for pricing financial and insurance risks. ASTIN Bulletin 32 (2), 213–234.

Willets, R., 1999. Mortality in the next millennium. Staple Inn Actuarial Society.

Willets, R., 2004. The cohort effect: Insights and explanations. British Actuarial Journal 10 (4), 833–877.

Wilmoth, J. R., 1990. Variation in vital rates by age, period and cohort. Sociological Methodology 20, 295–335.

Wilmoth, J. R., 1998. The future of human longevity: A demographer's perspective. Science 280 (5362), 395–397.

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.

Yang, S. S., Yue, J. C., Huang, H., 2010. Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. Insurance: Mathematics and Economics 46 (1), 254–270.

Yang, Y., Fu, W. J., Land, K. C., 2004. A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. Sociological Methodology 34, 75–110.

Zhou, R., Li, J. S.-H., Tan, K. S., 2015. Economic pricing of mortality-linked securities: A tatonnement approach. Journal of Risk and Insurance 82 (1), 65–95.

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.

Zhu, N., Bauer, D., 2011a. Applications of forward mortality factor models in life insurance practice. Geneva Papers on Risk and Insurance Issues and Practice 36, 567–594.

Zhu, N., Bauer, D., 2011b. Coherent modeling of the risk in mortality projections: A semi parametric approach. Tech. rep., Georgia State University.

Zhu, N., Bauer, D., 2014. A cautionary note on natural hedging of longevity risk. North American Actuarial Journal 18 (1), 104–115.