LONGEVITY AND STATISTICAL MODELLING

**YANG BOWEN
NANYANG BUSINESS SCHOOL**

2016

LONGEVITY AND STATISTICAL MODELLING

YANG BOWEN

Nanyang Business School

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2016

NANYANG TECHNOLOGICAL UNIVERSITY

Abstract

Nanyang Business School

Doctor of Philosophy

Longevity and Statistical Modelling

by YANG Bowen

This dissertation consists of two studies on the modelling aspects of mortality (or longevity). In the first paper, we examine cohort extensions of the Poisson common factor model for modelling mortality of both genders jointly. Several alternatives are specified and applied to datasets from five developed regions. We find that direct parameterisation of cohort effect could improve model fitting, reduce the need for additional period factors, and produce consistent mortality forecasts for females and males. Furthermore, we find that the cohort effect appears to be gender indifferent for the populations examined and has an interaction effect with age in certain cases.

The second paper explores the prediction error in mortality projection. This is important given the increasing longevity risk and the rising demand for longevity-linked products. Insofar, only parameter error and process error have been considered jointly while the issue of model error has been little studied. Here, we propose a method to account for process error, parameter error and model error in an integrated manner by modifying the semi-parametric bootstrapping technique. We apply the method to two datasets from the Continuous Mortality Investigation (CMI) and use the simulated scenarios to price the q -forward contracts with a risk-neutral approach. We find that model selection has a significant impact on the valuation results and thus it is crucial to incorporate model error in mortality projection.

The third part of the dissertation surveys the current landscape of the longevity market and discusses some open issues related to the pricing of longevity products in the context of the broader literature.

Preface

Part of the thesis have been published in:

Yang B., J. Li & U. Balasooriya (2016). Cohort extensions of the Poisson common factor model for modelling both genders jointly, *Scandinavian Actuarial Journal*, 2016(2), 93-112.

<http://dx.doi.org/10.1080/03461238.2014.908411>

Yang B., J. Li & U. Balasooriya (2015). Using bootstrapping to incorporate model error for risk-neutral pricing of longevity risk, *Insurance: Mathematics and Economics*, 62, 16-27.

<http://dx.doi.org/10.1016/j.insmatheco.2015.02.004>

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Prof Jackie Li and Prof Chen Zhanhui. Prof Li has introduced me to the field of actuarial research and provided me with ideas, insights and enormous guidance during my PhD years. Prof Chen has shown me the finance side of thinking and given me helpful advice and continued support in the later course of my research. Without them, I would not have accomplished this much.

I would also like to thank Prof Uditha Balasooriya for his kind advice and hearty encouragement. His positivity will continue to cheer me in future. My thanks also go to the other faculties from whom I have learnt and/or received help in one way or another. They are Prof Jan Kiviet, Prof Yao Shuntian, Prof Huang Weihong, Dr Michele Dacorogna, Dr Doug Andrews, Prof Tang Yang, Prof Angie Low, Prof Andreas Milidonis and Dr Yeoh Keng Leong. Just to name a few.

I'm very grateful as well to the division secretaries and all the PhD office staff and IT staff for their kind assistance throughout the years. Without their help, my learning journey would not be half as smooth.

I'm most fortunate to have a group of caring and helpful buddies: seniors, peers and juniors. Together, we have gone through the ups and downs of the PhD life. Kudos to us all. No matter where we go, we are comrades. Not to forget my table tennis and Taiji coaches and friends. The spirit of competitive sports and martial arts will be a treasure for me in life.

I'm truly thankful to all the examiners for their valuable time and efforts spent reviewing the thesis and for the constructive comments and feedbacks they provide.

Lastly, I want to thank my families and other friends. Your unconditional support means a lot to me. Thank you for bearing with my grumpiness, weakness and everything. I am here because of you.

Contents

Abstract	iii
Preface	v
Acknowledgements	vii
1 Introduction	1
2 Cohort Extensions of Poisson Common Factor Model	5
2.1 Poisson Common Factor Model	8
2.2 Cohort Extensions	11
2.3 Data Description	13
2.4 Result Analysis	14
2.4.1 Fitting Results	14
2.4.2 Forecasting Results	17
2.5 Out-of-Sample Testing	19
2.6 A Pricing Exercise	20
3 Model Error and Semi-Parametric Bootstrap	31
3.1 Mortality Models	34
3.1.1 LC Model	35
3.1.2 CBD Model with Curvature	35
3.2 Simulation with Semi-parametric Bootstrapping	36
3.2.1 Modified Simulation Procedures	38
3.3 Data Description	40
3.4 Simulation Results	41
3.4.1 Assured Lives	41
3.4.2 Pensioners	43
3.4.3 Accounting for Model Error	44
3.5 Risk-neutral Pricing via Maximum Entropy	46
3.5.1 Estimating the Risk-neutral Distribution	47
3.5.2 Results and Analysis	48
4 Longevity Risk and Current Developments	59
4.1 Pricing of Longevity Risk	62
4.2 Longevity and Macroeconomic Changes	64
4.3 Longevity and Financial Returns	66
Bibliography	73
A Estimation Procedure	83
B Other Residual Plots and Parameter Estimates	85

List of Figures

2.1	Life expectancy figures for females and males in Australia and the United States, 1950-2009.	22
2.2	Mortality indices for Australia and the United States, 1950-2009.	22
2.3	Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.	23
2.4	Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), France.	23
2.5	Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), West Germany.	24
2.6	Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), England and Wales.	24
2.7	Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), United States.	25
2.8	Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), Australia.	26
2.9	Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), France.	26
2.10	Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), West Germany.	26
2.11	Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), England and Wales.	27
2.12	Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), United States.	27
3.1	Mortality indices from LC (top left) and CBD (top right and bottom) models using assured lives data.	52
3.2	Log death rates from LC (top) and CBD (bottom) models using assured lives data.	52
3.3	Histograms of simulated present values of an annuity from LC (top) and CBD (bottom) models using assured lives data.	53
3.4	Mortality indices from LC (top left) and CBD (top right and bottom) models using pensioners data.	53
3.5	Log death rates from LC (top) and CBD (bottom) models using pensioners data.	54
3.6	Histograms of simulated present values of an annuity from LC (top) and CBD (bottom) models using pensioners data.	54
3.7	Histograms of simulated present values of an annuity given by the integrated approach using assured lives (top) and pensioners (bottom) data respectively.	55
3.8	The age profile of death rates at selected years for assured lives (left) and pensioners (right) data.	55

3.9	Forward prices of View the q -forwards with different maturity dates computed from LC model with process error only (dotted line), LC model with parameter and process error (short dashed line), CBD model with process error only (dot-dashed line), CBD model with parameter and process error (long dashed line) and our integrated approach (solid line) using assured lives (top) and pensioners (bottom) data. . . .	56
4.1	Best practice female life expectancy from 1840 to 2000.	71
B.1	Standardised residuals against age for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.	86
B.2	Standardised residuals against age for PLCM (left), PCFM (middle) and selected PCFC (right), France.	86
B.3	Standardised residuals against calendar year for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.	87
B.4	Standardised residuals against calendar year for PLCM (left), PCFM (middle) and selected PCFC (right), France.	87
B.5	Plots of the PCFC Model III parameter estimates, Australia.	88
B.6	Plots of the PCFC Model III parameter estimates, France.	89

List of Tables

2.1	BIC values for PCFM and PCFC, Australia.	28
2.2	BIC values for PCFM and PCFC, France.	28
2.3	BIC values for PCFM and PCFC, West Germany.	28
2.4	BIC values for PCFM and PCFC, England and Wales.	28
2.5	BIC values for PCFM and PCFC, United States.	28
2.6	Results of PCFM and selected PCFC for all five populations.	29
2.7	MAPE values of PCFC and PLCM forecasts.	29
2.8	Annuity factors obtained from PLCM and selected PCFC, United States.	29
3.1	Mean exposures and death rates over 1983–1994 and 1995–2006 at key ages for the assured lives, pensioners and England & Wales population data.	57
3.2	Proportion of simulated scenarios with a lower BIC value for each model.	57
4.1	Large longevity-linked transactions from 2010 to 2015. . . .	72

Chapter 1

Introduction

Mortality estimation and projection can be traced back to as early as the 17th century when modern life insurance was developed. For a long time, the techniques have been largely deterministic; not until the recent two decades, stochastic mortality modelling is gaining much attention in the research community. A vital impetus that fuelled this change was the ground-breaking work by Lee & Carter (1992). They demonstrated how to condense information from a large, cumbersome panel of mortality data into a set of three vectors that capture the essential features of the mortality profile of a population over time. The method is straightforward, parsimonious, and efficient, which explain its wide popularity in the literature. In addition, it has been used as a benchmark for US Census Bureau population forecasts and Social Security Administration retirement provisions (Hollmann et al., 2000; Lee and Miller, 2001).

Since Lee & Carter (1992), various extensions have been proposed to improve the modelling procedure and forecasting performance. Some experimented with incorporating higher order terms (Lee & Miller, 2001; Renshaw & Haberman 2003), using alternative estimation techniques (Delwarde et al., 2007; Koissi & Shapiro, 2006), or setting more realistic model assumptions (Brouhns et al., 2002). Other competing models that have been in the forefront are the age-period-cohort model (Bray, 2002) and the Cairns-Blake-Dowd model (Cairns et al., 2006). The latter is known for the superior performance in the older age ranges. For a comprehensive comparison study of the prevalent models, see Cairns et al. (2009).

The early focus was on the mortality of a single population. However, the mortality trends of small and data-scarce populations cannot be well captured and there exists severe parameter uncertainty (Andreev, 2002; Booth et al., 2006; Jarner et al., 2008). With this challenge in mind, Li et al. (2004) suggested borrowing information from certain reference populations with data of better quality. Jarner & Kryger (2011) proposed a more formal joint modelling framework that aims to track a small population's mortality with the help of a large reference population. They estimated the reference population's intensity surface with a frailty model and the spread between the two populations with a time-series model. In this way, the trend forecast would be more stable while the short- to medium-term fluctuations could be preserved.

In Chapter 2, we will provide a more detailed discussion of the existing works on and the merits of joint modelling. In particular, we will consider cohort extensions of one type of joint models, namely Poisson common factor model. A cohort refers to a group of people born in the same year. Thus, cohort effects can be taken synonymously as birth-year effects. Cohort effects have been widely studied in simpler uni-population applications (Currie, 2006; Renshaw & Haberman, 2006) but not as much in more sophisticated joint model set-ups. We explore in this direction and find that cohort effects play an important role in such instances as well.

You are probably aware that there is a basket of mortality models, some of which we have mentioned and will examine up close. In practice, given a data set, we often rely on some statistical measures to choose the "best" model for mortality projection. However, in doing so, we would run the risk of model misspecification. What if other models are more suitable but are discarded based on the available data? Traditional methods are able to

deal with parameter error and process error only. In Chapter 3, we propose to use semi-parametric bootstrap to incorporate model error in projection and valuation. We find that model selection has a significant impact on the valuation results and it is crucial to incorporate model error in mortality projection.

Thus far, the whole discussion is mostly about mortality modelling and forecasting. Observant readers may wonder why longevity in the title? The short answer is mortality and longevity are just two sides of the same coin, though depending on the viewpoint, one may be concerned with mortality risk more (less) than longevity risk.¹ Life insurers dislike mortality risk while pension funds are the opposite. As people live ever longer and experts repeatedly underestimate the survival rates at advanced ages (Mervyn King, 2004), the mass of discussion has shifted towards longevity risk, for good or bad, and for academics and practitioners alike. To be able to understand and manage this risk alone and in the presence of other risks, we certainly need better modelling and forecasting techniques, even more so than before. Chapter 4 sets on a short survey of the current developments and points out some issues for future research.

¹However, in some context, mortality risk may refer to specifically catastrophic risks, such as pandemics or natural disasters, that could bring losses of numerous lives.

Chapter 2

Cohort Extensions of Poisson Common Factor Model

Background

In the past decade or so, there is an emerging strand of literature that emphasises the interrelationship between multiple populations. Besides the two studies mentioned in the introduction, Plat (2009) used the notion of a stochastic portfolio experience mortality factor to address the issue of basis risk. Dowd et al. (2011) proposed a gravity model that specifies the stochastic evolution of the interaction between a large population and a small population. Cairns et al. (2011) adopted a Bayesian framework to model two related populations. Hyndman et al. (2013) modelled the product and ratio of mortality rates of two populations and produced coherent forecasts. On the other hand, there are some approaches that have a Lee & Carter (1992) origin. Li & Lee (2005) proposed an augmented common factor model, which incorporates a common factor and idiosyncratic components. The former captures the shared period effect while the latter targets the individual population characteristics. Li (2010) applied the augmented common factor model to the New Zealand population data. Building upon this concept, Li (2013) modified the error term assumption and the idiosyncratic components and applied it to the mortality of females and males in

the Australian population. The author showed that the ratio of the projected death rates at each age between females and males would approach a constant, mortality crossover can be precluded, and the cohort patterns in the residuals can be greatly reduced.

There are a number of compelling reasons for the joint modelling of multiple populations. First is biological consistency. It is not hard to imagine that populations that are geographically close to one another are likely to be influenced by a common set of fundamental drivers of mortality, such as economic conditions, the environment, and the state of medical technology. Examples of such related populations include populations of neighbouring countries, provincial versus national populations, and female versus male populations of the same region. For instance, Figure 2.1 shows that the female and male (period) life expectancy figures (collected from Human Mortality Database) move in tandem over time for Australia and the United States. Without joint modelling, mortality forecasts of these related populations could be divergent and nonsensical in the long run. Second is the need for hedging. As people live ever longer and the risk of longevity, i.e. the uncertainty in the rate of mortality improvement, exacerbates, insurance companies, private and state pension houses are looking for ways to off-load the risk to those who are ready to take it. The life market which is still at its nascent stage has seen an increasing use of mortality indices. Such indices have the advantage of being transparent and tractable and are therefore able to facilitate the structuring of financial contracts and consequently improve the liquidity of the life market. However, the indices are often derived from a certain reference population that is more often than not different from the actual insured portfolio. Therefore, co-modelling of the reference and insured populations is certainly necessary for quantifying the basis risk inherent in such transactions.

Moreover, the cohort effect has been a well-recognised feature in the mortality trends of many countries (MacMinn, 2003). Using insights gathered from epidemiology, economics and demography, Willets (2004) gave a detailed account of the cohort phenomenon observed in the mortality trends of the UK population. The main idea is that people born around the same time go through the same set of major life events, for example, recessions (economic), wars (political), famines (natural or man-made) or anti-smoking campaigns (social). If these major events occur at a critical stage in a person's development, then they are likely to leave deep imprints on the person's life, consequently affecting his or her later mortality. Thus, the cohort effect is a broad term that describes the aggregate effect on mortality of these events. Several quantitative analyses have been made to demonstrate that the cohort effect is indispensable in mortality modelling (Renshaw & Haberman, 2006; Richards et al., 2006). We believe that this feature is equally important in the joint modelling of multiple populations as in the modelling of a single population. Dowd et al. (2011) and Cairns et al. (2011) considered the cohort effect explicitly and co-modelled the cohort parameters (from the age-period-cohort model) of two populations with bivariate time series models. In this chapter, we consider a different approach and extend the models in Li & Lee (2005) and Li (2013) by incorporating a variety of cohort parameters. The results show that our proposed extensions could improve model fitting, reduce the need for additional period factors, and make consistent mortality forecasts between both genders.

The chapter is structured as follows. In Section 2.1, we first give a review of the Poisson common factor model in Li (2013). In Section 2.2, we elaborate on the extended models with cohort parameterisations. In Sections 2.4 to 2.5, we provide an analysis of the modelling results on datasets from five populations, including goodness-of-fit, forecasting performance,

and out-of-sample testing. In Section 2.6, we further illustrate the model performance in insurance pricing.

2.1 Poisson Common Factor Model

The force of mortality $\mu_{x,t,i}$ at age x in year t for gender i is assumed to be constant over an integer age-period interval, and thus the central death rate $m_{x,t,i}$ is equal to $\mu_{x,t,i}$. In the Poisson common factor model (PCFM) (Li, 2013), the number of deaths is modelled by a Poisson distribution:

$$D_{x,t,i} \sim Pn(E_{x,t,i}m_{x,t,i})$$

where $D_{x,t,i}$ is the death count and $E_{x,t,i}$ is the matching exposure to risk. The Poisson assumption is a more natural choice for representing the counting variable $D_{x,t,i}$ compared to the homoscedastic error term in Lee & Carter (1992) and Li & Lee (2005) though the death counts in many population data may be over-dispersed with the variance greater than the mean (Cairns et al., 2009). More importantly, as Li (2013) argued, it provides a rigorous framework for statistical analysis. Brouhns et al. (2002) and Renshaw & Haberman (2003) both applied this framework in their work.

The log death rate is in turn modelled as:

$$\ln m_{x,t,i} = a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} \quad (2.1)$$

where $a_{x,i}$ measures the overall age effect, K_t is the common mortality index that represents the shared period effect for both genders, B_x is the common age-to-period sensitivity that quantifies the changes in log mortality to changes in K_t for each age category. Together, $B_x K_t$ represents the common factor. Moreover, $b_{x,i,j} k_{t,i,j}$ is the j^{th} gender-specific factor, in which $k_{t,i,j}$ is the time component and $b_{x,i,j}$ is the age-to-period sensitivity. The higher-order gender-specific terms are part of the innovations in

Li (2013) that markedly improved the modelling results therein.

Following Booth et al. (2002) and Li (2013), the overall age effect and the common factor are estimated first via an iterative updating scheme (see Appendix A) that maximises the likelihood function, subject to the constraints $\sum_x B_x = 1$ and $\sum_t K_t = 0$ which ensure model identification. Then, taking $a_{x,i}$ and $B_x K_t$ as given, the same iterative method is used to quantify the gender-specific factors one by one.¹ Similar constraints are applied in the estimation of $b_{x,i,j}$ and $k_{t,i,j}$. Of course, the more parameters there are, the better the fit; however, there is always the danger of over-parameterisation. To strike a balance, the Bayesian Information Criterion (BIC) is used as an objective evaluation tool to determine the optimal number of gender-specific factors n . The BIC definition is taken as $-2\hat{l} + n_p \ln(n_d)$, where $\hat{l}$ is the computed log-likelihood, n_p the number of estimated parameters and n_d the number of observations used in estimation, and a lower BIC value is preferred. Considerations are also given to the patterns of the residual plots, the resulting trends of the parameters, and the amount of data being investigated.

Once the parameters are calibrated, the next step is to project the common mortality index K_t and the gender-specific time components $k_{t,i,j}$ for all j . The mortality index K_t is known for its highly linear nature² and is assumed to follow a random walk with drift:

¹Empirically, based on many trials, we find that running the entire updating scheme and computing all the parameters simultaneously does not necessarily lead to reasonable solutions. Often, the iterations appear to be trapped in some local maxima or fail to converge, despite using different initial values and constraints. The resulting parameter estimates are unstable and the likelihood values do vary. To circumvent these technical difficulties, we adopt the "step-wise" estimation approach as in Booth et al. (2002), Renshaw & Haberman (2006), and Li (2013). The final estimates might be slightly off the true global optimum, however, given that the parameters are estimated in order of importance, the deviations should be small. For more information on the fitting problems, see Currie (2014).

²A number of studies have pointed out the possible structural break that occurs around late 1960's and early 1970's for various OECD countries. See Li et al. (2011).

$$K_t = \delta + K_{t-1} + e_t \quad (2.2)$$

where δ is the drift term and e_t is a normally distributed random variable with mean 0 and variance σ^2 .

The time component $k_{t,i,j}$ is assumed to follow a weakly stationary $AR(p)$ model such that the gender-specific factors represent temporal deviation from the long-term trend. This helps ensure that the forecast ratios of male-to-female death rates will stay biologically reasonable and there will not be any mortality crossover in the projection. The $AR(p)$ model is specified as:

$$k_{t,i,j} = \alpha_{0,i,j} + \sum_{l=1}^p \alpha_{l,i,j} k_{t-l,i,j} + \epsilon_{t,i,j} \quad (2.3)$$

where $\alpha_{0,i,j}$ is a constant, $\alpha_{l,i,j}$'s are autoregressive coefficients, and $\epsilon_{t,i,j}$ is a normal error term. The future death rates can then be estimated accordingly.

Li (2013) applied the PCFM to an Australian dataset and produced better results compared to simply applying the Poisson Lee-Carter model (PLCM) to each gender separately. The forecast ratio of male-to-female death rates approaches a constant for all age groups, and the residual plots against cohort year display reasonable randomness. Although the additional gender-specific factors are able to capture some of the cohort patterns, they may not be the best candidates to do so. After all, the cohort effect and period effect, though both having a time dimension, are two different effects. In the next section, we will introduce some cohort extensions of the PCFM which incorporate different kinds of parameters that account directly for the cohort effect.

2.2 Cohort Extensions

Treating the PCFM as the baseline version (I), we suggest six competing models (II to VII) that characterise the cohort effect explicitly:

$$\begin{aligned}
 \text{(I)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} \\
 \text{(II)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + g_{t-x,i} \\
 \text{(III)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + g_{t-x} \\
 \text{(IV)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + c_{x,i} g_{t-x,i} \\
 \text{(V)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + c_x g_{t-x,i} \\
 \text{(VI)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + c_{x,i} g_{t-x} \\
 \text{(VII)} \quad \ln m_{x,t,i} &= a_{x,i} + B_x K_t + \sum_{j=1}^n b_{x,i,j} k_{t,i,j} + c_x g_{t-x}
 \end{aligned}$$

where the additional term in models II to VII represents the overall cohort effect. More specifically, g_{t-x} with or without subscript i denotes the “core” cohort effect (so named to distinguish from the overall cohort effect) while c_x with or without subscript i denotes the age-to-cohort sensitivity.

Note how these six models are interrelated. Model IV is the most generalised version amongst them. Models V and VI are immediately subsumed by model IV and have the following assumptions respectively: (1) the age-to-cohort sensitivity is the same for both genders and (2) the core cohort effect is the same for both genders. Model VII is a result of abiding by both of these restrictions. Models II and III are further contained by models V and VII respectively in that they have the extra assumption that the age-to-cohort sensitivity is equal to one across all ages.

In the literature, the cohort effect is either treated as an isolated effect or an interaction term with age. See Haberman & Renshaw (2011) for a review of various parametric mortality models, several of which include cohort parameters. Comparatively, our specifications here include the added dimension of gender, as well as several variants of the age interaction term. Examining these structures allows us to better understand the nature of cohort effect, for instance, whether it is gender-specific and/or age-specific.

Now we will elaborate on the estimation procedures. For each population considered, we estimate the PCFM with $n = 0, 1, \dots, 5$ additional factors. For each n , we condition on the computed values of $\hat{a}_{x,i}$, $\hat{B}_x \hat{K}_t$, and $\sum_{j=1}^n \hat{b}_{x,i,j} \hat{k}_{t,i,j}$ and estimate the cohort effect using the same iterative method (see Appendix A) that maximises the corresponding log likelihood function for each of the six specifications.

Accordingly, we have a set of 6×6 fitting results. To facilitate a fair comparison, we then calculate the BIC values for each of the 36 combinations. The following constraints are imposed on the cohort term for model identification and efficient convergence: $\sum_h g_h = 0$, where h takes the values of all possible cohort years and $\sum_x c_x = 1$. Note that if a gender subscript is present, the constraints will be applied on each gender separately. When estimating the core cohort vector, the first and last five cohort years are excluded from the estimation procedure so as to avoid high parameter uncertainty induced by a lack of data (e.g. Cairns et al., 2009).

Forecasting wise, the common mortality index is modelled by a random walk with drift as in equation 2.2. The time components of the gender-specific period factors are modelled by $AR(p)$ models as in equation 2.3. As the data used in the chapter only cover a period of forty years, the time

series of the time components are fairly short. Thus, lower order $AR(p)$ models are generally preferred (for most cases here, $p \leq 3$) unless there is strong evidence for the use of higher order autoregressive terms. Overall, the order p is determined by the autocorrelations of the residuals and the convergence of the forecast mortality ratios. The same treatment applies to the core cohort parameters. If the projected values of the fitted $AR(p)$ models do converge, the forecast male-to-female ratio of death rates at each age also tends to a constant. To illustrate, suppose the most recently observed death rates in year s are taken as the starting point of projection, the projected death rate (take model IV as an example) in year $t > s$ is therefore:

$$\tilde{m}_{x,t,i} = m_{s,t,i} \exp \left\{ \hat{B}_x(\tilde{K}_t - \hat{K}_s) + \sum_{j=1}^n \hat{b}_{x,i,j}(\tilde{k}_{t,i,j} - \hat{k}_{s,i,j}) + \hat{c}_{x,i}(\tilde{g}_{t-x,i} - \hat{g}_{s-x,i}) \right\} \quad (2.4)$$

Convergence of the $AR(p)$ models means that as t gets larger, $\tilde{k}_{t,i,j}$'s and $\tilde{g}_{t-x,i}$ in equation 2.4 will approach a constant, and as a result, the projected (log) ratio of death rates in equation 2.5 will converge. The same reasoning applies to the other five cohort models.

$$\begin{aligned} \frac{\tilde{m}_{x,t,1}}{\tilde{m}_{x,t,2}} &= \frac{m_{s,t,1}}{m_{s,t,2}} + \sum_{j=1}^n (\hat{b}_{x,1,j}(\tilde{k}_{t,1,j} - \hat{k}_{s,1,j}) - \hat{b}_{x,2,j}(\tilde{k}_{t,2,j} - \hat{k}_{s,2,j})) \\ &\quad + \hat{c}_{x,1}(\tilde{g}_{t-x,1} - \hat{g}_{s-x,1}) - \hat{c}_{x,2}(\tilde{g}_{t-x,2} - \hat{g}_{s-x,2}) \end{aligned} \quad (2.5)$$

2.3 Data Description

We apply the six alternatives of cohort extensions to datasets from five populations, namely Australia, France, West Germany, England and Wales, and the United States, obtained from the *Human Mortality Database*. The database profiles for an age range from 0 to 110+. However, the number of

deaths and exposures data are too volatile for more advanced ages, therefore the range from 0 up to 89 is used to allow for more precise estimation. A sample period from 1970 to 2009 is used for modelling purposes. The starting year of 1970 is chosen because of the documented "structural shift" that occurs around that time.³ For instance, Australia and the United States display visible structural changes in their mortality indices as shown in Figure 2.2. Choosing such a sample period helps ensure data relevance and enhance model performance. The caveat is that it might underestimate rare event probability when assessing the underlying variability. The ending year of 2009 is chosen as it is the latest year that all five populations have data available.

2.4 Result Analysis

2.4.1 Fitting Results

In this section, we present the modelling results for the five populations. We will analyse the results from two perspectives, namely the BIC values and the standardised residual plots. Table 2.1 shows the BIC values for the baseline PCFM (I) and the six cohort extensions (II to VII; denoted as PCFC) for the Australian dataset. Column I reproduces the results in Li (2013) that two gender-specific period factors (i.e. $n = 2$) is the optimal choice. On the whole, we observe that direct modelling of cohort effect reduces the need for additional gender-specific period factors: from models II to VII, all have the optimal n being less than two. Moreover, careful specification of the cohort effect helps improve the BIC value and in this case, model III with $n = 1$ is the best choice among the six alternatives. These observations suggest that direct cohort modelling would be a more

³See Booth et al. (2002) for a statistical measure to decide the optimal starting year for the Lee-Carter model. See Li et al. (2011) on the detection and implication of structural changes in the Lee-Carter mortality indices. If an earlier starting year is taken instead and the structural shifts are included, a more flexible model such as a piecewise linear curve (Li et al., 2011) may need to be used for the mortality index Kt .

efficient way in capturing the cohort effect.

For France, the BIC values are presented in Table 2.2. Again, we see a reduction in the optimal number of gender-specific period factors. In terms of the BIC value, all the cohort models outperform the baseline model substantially, with model VII being top in the race. Results of similar patterns are obtained for West Germany as shown in Table 2.3, except that the optimal n is reduced by two and the best cohort model is III. Slightly different results are obtained for England and Wales and the United States. As shown in Tables 2.4 and 2.5, when cohort parameters are included, the optimal number of period factors of the best model remains the same. This implies that the period factors have mostly been taking care of the period effect, and that the addition of cohort terms further boosts the model fit. Cairns et al. (2009) also found that inclusion of cohort parameters significantly reduces the BIC values for data of England and Wales and the United States.

Table 2.6 summarises the results of the PCFM and the selected PCFC for the five populations. The impact of incorporating cohort terms is apparent – the BIC value is lower for all cases and the optimal number of period factors is smaller for some cases. What is also worth mentioning here is that models III and VII seem to be the perennial winners. Both models have identical cohort effects for both genders, which suggest that the cohort effect is largely gender indifferent. Moreover, the age-to-cohort sensitivity is required for some populations but not for the others. Thus, whether the cohort effect is age-specific is highly dependent on the population of interest. Overall, the selected PCFC helps improve the BIC values in the range of a few hundreds to several thousands. It appears that the extent of improvement depends on how significant the underlying cohort effect is, which can be revealed by examining the residual plots.

To take a better look at how the cohort models improve the fitting results, we plot the standardised deviance residuals against cohort year for fitting the Poisson Lee-Carter model (PLCM) to each gender separately, the PCFM, and the selected PCFC side by side for comparison, as shown in Figure 2.3 to 2.7. Across all five populations, there is marked improvement in the randomness of the residuals from the left to the right side of the figure. The wavy patterns evident in the residual plots for the PLCM (left panel) are dampened significantly in the plots for the PCFM (middle panel) and are further reduced in the plots for the selected PCFC (right panel). Note that there may be over-dispersion in the data and the residuals are standardised with regard to the dispersion parameter (see Li 2013).

Zero in on each population. It can be seen that for Australia, the "second stage" improvement (i.e. from middle to right) is marginal, which corroborates the moderate reduction in the BIC value when the cohort effect is directly allowed for (Table 2.6). For the other four populations with large BIC reductions, the second stage improvement is more obvious. For France, West Germany, and England and Wales, the spikes or abnormal mortality associated with generations born around the time of Spanish flu in 1918 and WWII in the 1940's all fade into the background. Similar changes can be observed for the United States. These sudden cohort spikes are certainly hard to capture with the period factors that are designed to account for the period effect. This explains why the extended cohort models have superior fitting results.

We have also examined, though not fully reported, the residual plots against age and calendar year (see Appendix B). The selected PCFC performs well in these two aspects. We also provide in Appendix B the estimated parameters for the Australian and French datasets. It can be seen

that the common mortality index is highly linear while the other period and cohort series demonstrate some extent of autoregressive behaviour.

A final note is that the analysis above is heavily based on the BIC values.⁴ Although the BIC is the most commonly used criterion in the actuarial literature on mortality modelling, some may argue that its penalty on the excessive use of parameters is insufficient and there is a risk of overparameterisation. In this regard, an alternative is to use other information criteria with a greater emphasis on parameter parsimony. For example, McCullagh & Nelder (1989) discussed the criterion $Q = D + \alpha n_p \phi$, in which D is the deviance function, ϕ is the dispersion parameter and α ranges from 2 to 6, the choice of which depends on the level of penalty required. A higher level of penalty is expected to give rise to an optimal model with fewer parameters.

Although the fitting results of the PCFC are satisfactory in general, this does not necessarily mean that the proposed models are suitable for forecasting purposes. In the next few sections, we will analyse the forecast results of the selected PCFC.

2.4.2 Forecasting Results

As mentioned earlier, the common mortality index K_t is modelled with a random walk with constant drift. It is debatable if the linear trend in the mortality index k_t would extend into the future. Many tend to believe that there is an upper bound to the maximum human life span and the linear trend is likely to taper off (Fries, 1980; Hayflick, 1996 & 2000). However, with rapid advancement in biology and genetics research, some hold more

⁴As a robustness check, we have tried to compute the corresponding AIC values. It places lesser penalty on overfitting than BIC. Many countries end up with more gender-specific period factors. Nonetheless, the general results that including cohort terms reduces the need for more period factors and improves model fit remain.

optimistic views and bet on trend continuation (Manton et al., 1997; Ricklefs & Finch, 1995; Guarente, 2000; Christensen et al., 2009). Aware of the debates, we are cautious not to overextrapolate the trend of k_t forward. The forecast horizon is set to be no greater than the sample period. The projected linear trend is merely best estimate. In Chapter 3, we further account for various errors in our projection to be more conservative. The time components of the gender-specific period factors $k_{t,i,j}$'s are modelled as weakly stationary $AR(p)$ time series. The cohort series are assumed to be autoregressive as well. The forecast ratio of male-to-female death rates at each age will then converge to a constant in the long run and tends to stay within a reasonable range in accordance with empirical observations.

This property is well reflected in the forecast ratio charts, shown in Figure 2.8 to 2.12. In each figure, the left chart is the projection of the PLCM which is applied to females and males separately while the plot on the right is the projection of the selected PCFC. Consider Australia as an example (Figure 2.8). The forecast ratio by the PLCM for age group 30-49 is trending rapidly upward and that for age group 50-69 is trending downward drastically. If a longer projection horizon is used, these two ratios could go as high as above three and as low as below one respectively, which are biologically unsound. It is difficult to justify such long-term divergence behaviour between female and male mortality. Whereas if the cohort model is used, we can see that the trends are moderated and plateau to a constant level. Similar phenomenon can be observed for France age groups 0-9, 10-29 and 70-89, for West Germany age groups 0-9, 70-89, for England and Wales age groups 10-29, 30-49 and 50-69, and for the United States age group 50-69. Overall, the forecast ratios from the PCFC are broadly in line with past observations and are rather consistent across different age groups.

2.5 Out-of-Sample Testing

In this section, we conduct out-of-sample testing, commonly known as back-testing, to examine how well the PCFC performs in terms of forecasting. We divide the 40 years of data into two halves – the first 20 years for estimation and the remaining 20 years for forecasting and back-testing. Specifically, for each population, we use the data from 1970 to 1989 to estimate the parameters and select the best PCFC based on the BIC.⁵ Then we forecast the mortality and mortality ratio, from year 1990 to 2009. The accuracy of the forecast mortality and mortality ratio is evaluated against the actual values by computing the mean absolute percentage error (MAPE). We also calculate the MAPE values for applying the PLCM to each gender separately and provide them alongside for comparison.

The out-of-sample testing results are tabulated below in Table 2.7. The left panel shows the MAPE values of forecast death rates and the right panel shows the MAPE values of forecast ratios of male-to-female death rates. In each panel, the first and second columns display the MAPE values for the PCFC and the PLCM respectively, and the third column displays the percentage difference between the two values, i.e. how much better the PCFC is when compared to the PLCM.

As shown in Table 2.7, the MAPE values of forecast death rates for the PCFC range from 9.3% to 17.6%, with the United States and West Germany on the lower (better) end and Australia on the higher (poorer) end. For all five populations, the selected PCFC delivers a better MAPE value than the PLCM, though to varying degrees. This demonstrates the ability of our cohort models to produce fairly accurate forecasts. If we turn to the right panel, that is the MAPE values of forecast ratios of male-to-female

⁵With a shorter estimation period, the optimal n may reduce, and the best cohort model switches from VII to III for France but remains the same for the other populations.

death rates, we can see that for the PCFC, the errors mostly centre around 11%. Only Australia has a rather high MAPE value of 20%. For four out of the five populations, the selected PCFC outperforms the PLCM by a large margin of about 3% or more. In the case of England and Wales, the PCFC still wins by a small margin of 0.3%. This reaffirms our results in Section 6 that the PCFC produces more stable mortality forecasts between both genders. On the other hand, without joint modelling of both genders, the PLCM could produce forecasts of mortality ratios that deviate more from the actual values. Overall, the out-of-sample testing gives us another testimony to the accuracy and consistency of the extended cohort models.

2.6 A Pricing Exercise

Joint modelling of both genders can have an impact on pricing certain insurance products. As an illustration, we use the forecast death rates for the United States to calculate the premium at the beginning of calendar year 2010 for an annuity which pays \$1 at the end of each year for a term of 20 or 25 years. Assume that the person is 60 or 65 years old at the start and the annual interest rate is 3% flat. The calculation results are shown in Table 2.8. We can see that the difference in the premiums between a female and a male is smaller for the PLCM than for the PCFC. This is because with the PLCM, the mortality ratio for age group 50-69 is projected to be trending downward linearly, whereas for the PCFC, the ratio approaches a constant (Figure 2.12). The latter case seems to be more realistic, as the mortality ratio for age group 50-69 levelled off in the 10 years after around 2000 and it is undesirable to project a persistent downward trend which is incompatible with the other age groups. The implication is that the difference between female and male premiums with the PLCM is potentially underestimated by 7 to 14 cents, and effectively, the male annuitants would be cross-subsidising the female annuitants in the market. The cross-subsidisation would occur in an opposite way if we consider a whole of

life or term life insurance instead.

Summary

We have shown that the extended cohort models PCFC could improve model fitting measured by the BIC, MAPE, and the randomness in the residual plots. It reduces the need for additional period factors in the original PCFM, which is part of the reasons that account for the satisfactory modelling results. On top of that, it produces consistent mortality forecasts between females and males in a population, which could translate to more reasonable pricing of insurance products. Another interesting finding is that the BIC values do not support the use of separate cohort parameters for each gender and the cohort effect appears to be gender-indifferent for the populations under study. Further research on this matter may be carried out for other populations and time periods.

Future work includes testing different multivariate time series models on the period and cohort parameters, with, for example, a VARIMA process (Chan et al., 2014), as it is important to evaluate the underlying uncertainty of future outcomes. It would also be useful to apply the PCFC to industry and population data and measure the basis risk of hedging an insurance portfolio with standardised mortality derivatives.

The number of deaths has been assumed to obey the Poisson distribution. However, there could be the problem of overdispersion due to, for instance, heterogeneity in the population of a given age-period-gender group or small exposure at advanced ages. An alternative we could investigate in future is to consider the negative binomial distribution, which has an additional parameter to control for overdispersion and which in the limit converges to the Poisson distribution.

FIGURE 2.1: Life expectancy figures for females and males in Australia and the United States, 1950-2009.

FIGURE 2.2: Mortality indices for Australia and the United States, 1950-2009.

FIGURE 2.3: Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.

FIGURE 2.4: Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), France.

FIGURE 2.5: Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), West Germany.

FIGURE 2.6: Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), England and Wales.

FIGURE 2.7: Standardised residuals against year of birth for PLCM (left), PCFM (middle) and selected PCFC (right), United States.

FIGURE 2.8: Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), Australia.

FIGURE 2.9: Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), France.

FIGURE 2.10: Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), West Germany.

FIGURE 2.11: Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), England and Wales.

FIGURE 2.12: Forecast ratios of male-to-female death rates, PLCM (left) and selected PCFC (right), United States.

TABLE 2.1: BIC values for PCFM and PCFC, Australia.

n	I	II	III	IV	V	VI	VII
0	71,505	69,746	70,119	70,165	69,995	69,587	70,065
1	69,705	69,728	69,126	70,587	70,181	70,038	69,605
2	69,625	70,542	69,710	71,664	71,096	70,890	70,274
3	70,882	71,933	71,054	73,164	72,514	72,326	71,645
4	72,457	73,677	72,790	74,879	74,244	74,049	73,401
5	74,156	75,448	74,551	76,692	76,010	75,876	75,171

TABLE 2.2: BIC values for PCFM and PCFC, France.

n	I	II	III	IV	V	VI	VII
0	113,853	93,559	102,017	90,495	90,067	96,914	99,813
1	101,704	90,144	92,198	88,608	88,884	89,876	90,599
2	92,611	85,512	84,978	85,138	84,732	84,818	84,335
3	90,522	85,163	84,418	85,116	84,437	84,435	83,760
4	90,292	85,376	84,673	86,196	85,483	85,524	84,899
5	91,100	87,018	86,155	88,138	87,481	87,519	86,559

TABLE 2.3: BIC values for PCFM and PCFC, West Germany.

n	I	II	III	IV	V	VI	VII
0	113,901	90,057	96,744	88,432	88,888	96,201	95,877
1	103,163	86,010	86,715	86,271	85,895	87,402	86,896
2	99,485	85,285	85,132	86,032	85,293	86,018	85,434
3	98,059	85,906	85,192	86,654	86,047	86,357	85,714
4	97,237	87,022	86,105	87,776	87,183	87,269	86,642
5	97,574	89,271	88,351	90,041	89,452	89,462	88,851

TABLE 2.4: BIC values for PCFM and PCFC, England and Wales.

n	I	II	III	IV	V	VI	VII
0	114,512	91,577	103,429	88,373	87,952	96,467	100,239
1	96,706	86,726	88,212	85,651	85,057	87,597	87,323
2	88,189	82,310	82,099	82,662	82,248	82,516	82,053
3	86,731	82,334	81,825	83,100	82,552	82,799	82,219
4	87,142	83,256	82,687	84,209	83,876	83,872	83,207
5	88,462	85,280	84,654	86,213	86,012	85,846	85,190

TABLE 2.5: BIC values for PCFM and PCFC, United States.

n	I	II	III	IV	V	VI	VII
0	294,121	201,884	263,731	174,812	183,702	183,934	250,496
1	199,806	169,279	178,568	150,014	154,101	154,371	163,990
2	145,755	135,042	136,989	129,568	131,813	133,093	133,616
3	117,020	109,465	109,881	108,064	109,302	109,405	109,409
4	110,376	104,851	104,508	104,552	104,037	104,483	103,846
5	108,797	103,660	103,157	104,096	103,471	103,793	103,091

Note: More than five period factors have been tested to determine the optimal n , and $n = 5$ is indeed the best. However, for a consistent presentation format, results for only up to five factors are shown here.

TABLE 2.6: Results of PCFM and selected PCFC for all five populations.

Population	PCFM		Model	PCFC	
	n	BIC		n	BIC
Australia	2	69,625	III	1	69,126
France	4	90,292	VII	3	83,760
West Germany	4	97,237	III	2	85,132
England & Wales	3	86,731	III	3	81,825
United States	5	108,797	VII	5	103,091

TABLE 2.7: MAPE values of PCFC and PLCM forecasts.

Population	Death rates			Male-to-female ratios		
	PCFC	PLCM	$\Delta(\%)$	PCFC	PLCM	$\Delta(\%)$
Australia	0.1764	0.1840	0.76	0.2006	0.2505	4.99
France	0.1292	0.1409	1.17	0.1179	0.1497	3.18
West Germany	0.0972	0.1132	1.60	0.1143	0.1501	3.57
England & Wales	0.1265	0.1299	0.34	0.1116	0.1154	0.38
United States	0.0927	0.0993	0.66	0.1113	0.1412	2.99

TABLE 2.8: Annuity factors obtained from PLCM and selected PCFC, United States.

Age	Term	PLCM			PCFC		
		Female	Male	F-M	Female	Male	F-M
60	20	14.12	13.61	0.51	14.15	13.57	0.58
65	20	13.50	12.83	0.67	13.58	12.83	0.75
60	25	15.98	15.24	0.74	16.04	15.17	0.87
65	25	14.99	14.03	0.95	15.09	14.00	1.09

Chapter 3

Model Error and Semi-Parametric Bootstrap

Background

Since the pioneering work by Lee and Carter (1992), a large number of discrete-time stochastic mortality models have been proposed. Roughly speaking, they can be divided into several broad, though not necessarily distinct, categories: the Lee-Carter (LC) model and its variants, the age-period-cohort (APC) model (Bray, 2002; Currie, 2006), the Cairns-Blake-Dowd (CBD) model and its extensions (Cairns et al., 2006), and the p-spline model (Currie et al., 2004). New models and extensions are often proposed in an attempt to improve the modelling and forecasting performance of the earlier versions.

With a variety of tools at our disposal, it is natural that we ask the question which model is the best and should be used in practice. Unfortunately, there is no short answer to this question. It is often the case that the choice of model depends on the data available as well as the user's preference. Given the data and certain qualitative requirements, it might still be difficult to pick the single "best" model from numerous potential candidates.

Several comparative studies have been done to examine the relative

performance and robustness of these models. They are generally based on some statistical criteria for model fitting within sample and for forecasting accuracy out of sample. For example, Cairns et al. (2009) examined the goodness-of-fit as well as the robustness of parameter estimates relative to the estimation period. Renshaw and Haberman (2011) employed similar criteria in their study with more competing models. Dowd et al. (2010) adopted a series of backtesting procedures that rigorously evaluate the model forecasts against the realised mortality. They have also suggested a number of qualitative criteria for evaluating the models. Cairns et al. (2009) put forward a set of these conditions such as parsimony, transparency, ease of implementation, ability to generate sample paths, and so on.

The consensus is probably that no single model stands out as the best under all circumstances. The projections and simulations from different models can be quite different, which could lead to considerable model error, i.e. error due to model misspecification. This problem becomes a real concern in pricing longevity-linked instruments, which annuity providers and pension fund sponsors are eager to use so as to transfer undesirable longevity risks on their book. If model error is understated, the capital market investors might be unwilling to accept the deal in the first place. On the other hand, the annuity providers and pension fund sponsors certainly do not want to overpay for the deal. Hence, it is important to strike a balance and quantify the real impact of model error on the transaction prices. As noted in QIS5 Technical Specifications (2010) for Solvency II, “the nature of the risks underlying the insurance contracts could be described by the probability distribution of the future cash flows arising from the contracts”, and process error, parameter error and model error should all be considered.

Risk-neutral valuation generally involves incorporating risk adjustment into the real-world probability distribution by converting it to a risk-neutral distribution using market prices, such that the expected return in a risk-neutral world is equal to the risk-free rate. In the literature, there are a number of simulation methods that can be used to generate the real-world distribution, taking into account process error and parameter error, e.g. Lee and Carter (1992), Li et al. (2004), Brouhns et al. (2002b), Brouhns et al. (2005), Czado et al. (2005), and Koissi et al. (2006). Nevertheless, not much effort has been made to directly assess the potential impact of model misspecification and examine the three sources of uncertainty in an integrated manner. In this chapter, we thus attempt to address the concern of model error. We propose to modify the semi-parametric bootstrap method in Brouhns et al. (2005) in order to simulate and integrate the projections from two or more mortality models. Using our proposed approach, uncertainty of model selection can be allowed for, in addition to random fluctuations and uncertainty in parameter estimation. The simulated outputs can then be used to estimate the “likelihood” of each model, as well as to price longevity-linked instruments.

In Section 3.1, we will first describe the two mortality models we consider in this study. In Section 3.2, we will introduce our modification to the semi-parametric bootstrapping that has been used to assess only process error and parameter error. We will then apply it to two data sets from the Continuous Mortality Investigation (CMI) and analyse the simulation results in 3.4. In Section 3.5, we will use the simulated scenarios in a pricing exercise to examine the incremental impact of incorporating model error. Finally, in the Summary, we will conclude and provide suggestions for future research.

3.1 Mortality Models

For the purpose of this chapter, we consider two representative mortality models, namely the Poisson Lee-Carter (PLC) model and the Cairns-Blake-Dowd (CBD) model with a curvature term. For both models, we describe the number of deaths as a counting variable with a Poisson distribution. Such specification is more intuitive and provides a sound statistical framework for estimation and analysis. Moreover, the curvature term added to the CBD model can better depict the non-linearity in the effect of age on (log) mortality.

We choose these two models for our study because they belong to the two most popular branches of mortality models in the literature, and can capture the essential mortality features of the assured lives and pensioners data used here. Moreover, they are different enough for us to examine the potential extent of model error.

These two models have various extensions; for example, see Cairns et al. (2009). There are also other approaches that do not belong to these two branches, such as projecting the parameters of a mortality curve (McNown and Rogers, 1989) and the age-period-cohort model (Bray, 2002). However, we restrict our attention to the two models so that we can concentrate on how to factor model error in mortality projection. Our method can be extended to incorporate other or more than two mortality models.

In what follows, we will briefly review the LC and CBD models and explain the fitting and projection procedures. For more details, refer to Brouhns et al. (2002a) and Tan et al. (2014).

3.1.1 LC Model

The Poisson LC model is a special case of the PCFM model, with the gender subscript as well as the idiosyncratic components suppressed, i.e. we assume there is no difference in the mortality of the two genders. As a result, only the age-specific factor and the common factor count:

$$\ln m_{x,t} = a_x + b_x k_t \quad (3.1)$$

The relevant parameter constraints apply. As before, maximum likelihood is used to estimate the parameters, and the estimated time series of k_t is then modelled as a random walk with drift for projection purposes.

3.1.2 CBD Model with Curvature

In the CBD model with an additional curvature term (Cairns et al. 2009; Tan et al., 2014), the number of deaths is also distributed as a Poisson variable. However, instead of modelling the log death rate, this model deals with the logit death probability $q_{x,t}$ as below,¹

$$\ln \left\{ \frac{q_{x,t}}{1 - q_{x,t}} \right\} = k_{1t} + k_{2t}(x - \bar{x}) + k_{3t}((x - \bar{x})^2 - \sigma_x^2) \quad (3.2)$$

where $\bar{x}$ is the average age in the sample and σ_x^2 is the mean squared age deviance. The term k_{1t} refers to the general level of mortality, k_{2t} describes the age differentials, and k_{3t} depicts the curvature of the logit mortality curve. Maximum likelihood is again used to estimate the parameters, and no constraints are required. The estimates for a particular period t are only influenced by the data from that period, thus the times series of k_{1t} , k_{2t} , k_{3t} are said to be “time-invariant”, a property that makes them potential candidates for building mortality indices to be traded in the financial market

¹If $\mu_{x,t}$ is assumed to be constant in each age-period cell, $q_{x,t}$ is related to $\mu_{x,t}$ through the equation $q_{x,t} = 1 - \exp(-\mu_{x,t})$.

(Chan et al., 2014). For projection, these three time series can be modelled as a trivariate random walk with drift.

3.2 Simulation with Semi-parametric Bootstrapping

There are several simulation methods for incorporating process error (random fluctuations) and parameter error (uncertainty in parameter estimation). In Lee and Carter (1992), the authors mainly accounted for the random error of projecting the mortality index. Li et al. (2004) took a step further by considering the estimation error of the drift term of the mortality index.

However, the extant mortality models are often characterized by a large number of parameters. Focusing on the sampling error of only one or two parameters may not be adequate. In order to capture parameter error in its entirety, Brouhns et al. (2002b) adopted a parametric Monte Carlo simulation approach and generated samples of the model parameters from a multivariate normal distribution. In a follow-up study, Brouhns et al. (2005) examined a semi-parametric bootstrap procedure that simulates death counts from a Poisson distribution with the mean equal to the observed number of deaths. In contrast, Renshaw and Haberman (2008) used the fitted number of deaths as the mean of the Poisson distribution. Following the line of thought but in a different fashion, Koissi et al. (2006) bootstrapped from the residuals of a fitted LC model.

An alternative simulation method was proposed by Czado et al. (2005) and Kogure et al. (2009). This method is based on their Bayesian adaptations of the Lee-Carter model. They derived posterior distributions for the parameters and used Markov chain Monte Carlo (MCMC) simulation to generate sample paths and factor in parameter error.

Different simulation procedures have their own strengths and weaknesses. Brouhns et al. (2005) compared the parametric and semi-parametric procedures using the Belgian population data (1950-2000) and found that they give rise to largely comparable sampling results. Renshaw and Haberman (2008) studied these two simulation methods together with residual bootstrapping. They found that the parametric Monte Carlo approach would lead to widely differing prediction intervals when different constraints are set. Using the New Zealand mortality data (1948-2008), Li (2014) performed a quantitative comparison of all the six simulation strategies mentioned above. The author concluded that the sampling results from the parametric, semi-parametric, residual bootstrap, and Bayesian MCMC methods are fairly close to one another. The author also commented that the Lee and Carter (1992) and Li et al. (2004) approaches are straightforward to implement, whereas the other four methods offer a more sophisticated mechanism to incorporate different errors.

Despite these interesting proposals and findings, there is one important risk component missing: model error (i.e. error due to model misspecification). Relatively little attention has been paid to this important problem. Cairns (2000, 2014) suggested a Bayesian framework to set a prior probability for each model considered. Denuit (2009) proposed using a model selection criterion to assign a weight to each model. In the literature, researchers often compared different models based on certain statistical criteria and qualitative features, without incorporating directly the potential effect of model misspecification. Thus, we attempt to supplement this knowledge gap. We propose to modify the semi-parametric bootstrap method in Brouhns et al. (2005) for testing two or more models simultaneously. While we demonstrate this modified approach with only the LC and CBD models and a single statistical criterion, the framework is general in nature and can readily be adapted to many other choices. Next,

we will proceed to describe the simulation steps in detail.

3.2.1 Modified Simulation Procedures

Following Brouhns et al. (2005), we first bootstrap n samples $\{d_{x,t}^{(j)}\}_{j=1}^n$ from the Poisson distribution with the mean equal to $d_{x,t}$, the observed number of deaths², for each age and period. That is, each sample consists of a panel of pseudo death counts of the population.

Suppose there are m competing models under consideration. For the j^{th} sample, we fit each model in turn to the pseudo data and estimate the parameters. Then based on some pre-specified model selection criteria (quantitative or qualitative), the “best” model is chosen for this particular sample. The time series of the model’s parameter estimates are projected into the future, taking into account the stochastic nature of the time series process. Consequently, we obtain simulated future mortality rates using the estimated parameters and the projected time series from the best model. The above steps are repeated to produce a total of n scenarios.³ In this way, all of process error, parameter error, and model error would be incorporated in the simulations.⁴

²Using the observed number of deaths as the mean for the Poisson distribution in Brouhns et al. (2005) is dissimilar to the common bootstrapping approaches in the literature, which use the fitted value based on a particular model, as the mean. Here we choose to adopt the former method as we are interested in producing “model-free” pseudo death counts such that all the competing candidate models could be evaluated on a level playing field.

³As noted in Chernick (1999), the bootstrap method does not require any analytical expression for the estimator. It can be used as long as “there is a computational method for deriving the estimator”. Most simulation studies suggest that the bootstrap method can safely be adapted to a large number of problems where there is no strong theoretical justification yet.

⁴When the data set or portfolio size is small, for instance, in the case of an individual insurer, there may be a need to further consider sampling error, i.e. the observed number of deaths deviating from the expected value by chance. This sampling error can be accounted for by using, for instance, a binomial distribution. Since industry aggregate data are used here, sampling error would not be sizable and is ignored for convenience.

For demonstration, as discussed in Section 2, we set $m = 2$ and consider the LC model and CBD model with a curvature term. After examining a number of alternatives, we find that these two options give rise to reasonably random residuals over age, calendar year, and cohort year for our assured lives and pensioners data, and produce parameter series which can readily be projected with a random walk with drift. Moreover, these two candidates belong to two major and distinct branches of models in the literature, which in our view offer sufficient diversity for the model specification process. We estimate the parameters $\{a_x^{(j)}, b_x^{(j)}, k_t^{(j)}\}$ and $\{k_{2t}^{(j)}, k_{1t}^{(j)}, k_{3t}^{(j)}\}$ for the j^{th} sample and project $\{k_t^{(j)}\}$ by a random walk with drift and $\{k_{2t}^{(j)}, k_{1t}^{(j)}, k_{3t}^{(j)}\}$ by a trivariate random walk with drift.

With regard to the model selection criteria, for illustration, we adopt the Bayesian Information Criterion (BIC) as the sole indicator. This statistical criterion is probably the most popular choice in mortality projection research. However, other statistical criteria can also be used, such as the Akaike Information Criteria (AIC) and various statistical tests for the randomness of residuals. Besides quantitative conditions, some qualitative requirements may also be imposed. For example, for a particular pseudo sample, a candidate model can be rejected if it produces biologically unreasonable projection such as having the projected death rates decreasing over age for a certain period.

Alternatively, a candidate model can be chosen over the others if it produces the least varying (i.e. robust) parameter estimates when fitted to different subperiods of the pseudo sample. Qualitative criteria of these kinds can be translated into executable programs, though the computation time will have to increase. See Cairns et al. (2009, 2011) for more discussions on desirable model properties.

When there are two or more criteria in place, a weight or preference could be assigned to each criterion so that there is an overall score for ranking the m models in each simulation step. In practice, one may wish to incorporate more than two models and/or multiple selection criteria, which will likely increase the running time significantly. This problem, however, will become less of an issue as computer technology advances over time.

In the following, we simulate $n = 10,000$ scenarios of future death rates. The simulated values can be used to form the distributions of key mortality indicators, such as central death rates, life expectancy, and present value of an annuity.

3.3 Data Description

We conduct simulation on two sets of data. One is for assured lives and the other for pensioners. Both are obtained from the *Continuous Mortality Investigation* (CMI), a task force that collates and analyses data from UK life assurance companies and actuarial consultancies. The characteristics of the two data sets would be different due to reasons such as the degree of adverse selection, existence of group insurance, underwriting requirements, socio-economic class etc. Having two sets of data would thus allow us to see how the two mortality models perform on their own and collectively under different situations.

The death and exposure data set for assured lives has a longer history (1947-2006) than that for pensioners (1983-2006). To enable a meaningful comparison across the data sets, we limit our study to the period from 1983 to 2006. This period is also more relevant as it avoids the more distant and obsolete trends in mortality. Many studies have found structural breaks in the Lee-Carter mortality index around the 1970s for various countries, suggesting that the fundamental drivers of mortality may have changed

after that point (Renshaw and Haberman, 2003; Li, 2013; Li et al., 2011). As for age, we choose the range from 65 to 95, as we focus on longevity risk at retirement ages. Data beyond age 95 is scarce and unsuitable for use in our modelling and simulations.

Table 3.1 provides some summary statistics for the CMI data sets alongside the England and Wales population data from the *Human Mortality Database* for comparison. The mean exposures and death rates over the subperiods at key ages are computed for each data set. The CMI data sets have much smaller exposures than the population data. Between the two, the assured lives data set has larger exposures at younger ages than the pensioners data set but smaller exposures at older ages. In terms of death rates, the pensioners are comparable with the population. The assured lives, on the other hand, seem to have much lower rates. Over time, we see shrinking exposures for the younger ages but marked increase for the older ones. In addition, death rates at all ages are decreasing, especially for early retirement years.

3.4 Simulation Results

3.4.1 Assured Lives

We simulate 10,000 sample paths based on the data set for assured lives. Then we fit each of the two models to this set of pseudo data and project the estimated mortality indices. We first use the original Brouhns et al. (2005) approach for each model separately, without including model error. Figure 3.1 plots the computed mortality indices from 1983 to 2006, as well as the projected mean values and 95% prediction intervals for the next 30 years. We can see that the declining trend in k_t (from LC) is fairly similar to that in k_{1t} (from CBD with curvature). This is not surprising as both

are first-order measures of mortality improvement. Both are downward-sloping with some increase in variability in recent years, showing the persistent trend in mortality improvement and perhaps increasing uncertainty in its development.

Moreover, k_{2t} is positive and slightly upward-sloping, which means that mortality improvement occurs at a faster pace at the younger end of the sample age range. The decreasing, negative values of k_{3t} suggest that the logit mortality curve deviates from linearity over time, with the middle part of the age range bulging upward. See Tan et al. (2014) for more detailed descriptions of these mortality indices.

Figure 3.2 plots the observed and projected values of log death rates at ages 65 and 80. The top two charts are produced from the LC model while the bottom two from the CBD model with curvature. It is interesting to see that the LC model predicts faster mortality improvement at age 65 and its prediction interval is smaller than that by the CBD model. Comparatively, at age 80, the differences between the predictions from the two models are less obvious. These results highlight the fact that different models can produce varying results at different ages and points of time, and the issue of model misspecification looms large and should be properly assessed.

To insurance companies and pension plans, another important figure is the present value of an annuity. We consider a simple annuity that pays one pound at the end of each year until the person dies or reaches age 95, whichever is earlier. The interest rate is taken as 3% flat. Figure 3.3 shows the histograms of the simulated present values for a person aged 65 and 80 respectively at the start of 2007. Again the top charts are produced from the LC model and the bottom from the CBD model with curvature. We find that the sample mean values are largely the same between the two models.

For age 65, the mean values are 15.52 (LC) and 15.61 (CBD). For age 80, the mean estimates are 7.92 and 7.99 respectively. However, it can be seen that the LC model leads to lower variability in the simulated distributions. The 95% prediction interval for age 65 given by the LC model is (14.80, 16.24) and that by the CBD model is (14.67, 16.49). For age 80, the difference is more apparent. The LC model gives an interval of (7.52, 8.36) while the CBD model gives (7.27, 8.69). This difference in variability would have a significant impact on risk-neutral pricing, as will be discussed in Section 4.

3.4.2 Pensioners

We carry out the same simulation exercise for the data set on pensioners. In Figure 3.4, we see similar predictions in the trends of mortality indices as in Fig. 1. The parameter series k_t and k_{1t} decrease over time, k_{2t} increases, while k_{3t} does not change much.

For log death rates (Figure 3.5), the two models produce similar mean values but the LC model leads to lower variability in the projection, which is comparable to the earlier results for assured lives.

The histograms of the simulated present values of an annuity for age 65 and age 80 are shown in Figure 3.6. The variability in the projection based on the LC model is again lower than that on the CBD model. For age 65, the 95% prediction intervals are (13.91, 15.45) by the LC model and (13.93, 15.59) by the CBD model. For age 80, the intervals are (6.72, 7.39) by the LC model and (6.56, 7.65) by the CBD model. As noted above, this disparity in sampling distributions would affect the risk-neutral valuation of longevity risk.

3.4.3 Accounting for Model Error

Now we are ready to introduce model error into our simulations. As emphasised in Section 3, the same set of pseudo data is fed into the two mortality models for estimation and projection. In particular, for the j^{th} simulated sample path, we select the model that gives a better fit according to the BIC. Then we use the chosen model to generate future values for this particular sample path. We repeat this step for all samples ($j = 1, 2, \dots, n$) and obtain a total of $n = 10,000$ scenarios. In this way, the uncertainty of model selection is taken into account.

Some may argue that there are other measures to evaluate the goodness-of-fit and the BIC may not be the best one. Indeed, the main difference among the usual selection criteria is often the penalty placed on over-parameterization. It may be useful to test the sensitivity of the results against different criteria. However, for this study, we consider the BIC only and we acknowledge this limitation therein.

Denuit (2009) proposed a method to handle model risk. The author considered K prediction models and denoted them as $M_1, M_2, \dots, M_K$. Each model would be assigned a weight for a given data set D . To compute the weight for M_k , he suggested using the ratio $\frac{\exp(-I_k/2)}{\sum_{l=1}^K \exp(-I_l/2)}$, where I_k is the value of a chosen information criterion for M_k . The forecast value of any life quantity would be the weighted forecast of the K models. Our method is similar in that technically, all the candidate models are weighted. However, how exactly we assign the weightings is different – we base the weights on 10,000 simulated samples, while Denuit (2009) used the real data set directly.

In our simulations, the weights (i.e. proportions with a lower BIC value) of the two models for each data set are given in Table 3.2. It can

be seen that the LC model appears to be the more suitable option for the assured lives while the CBD model with curvature the better choice for the pensioners.

In Figure 3.7, we provide the histograms of the present values of an annuity calculated using the integrated approach. As the weightings would have suggested, the histograms for the assured lives are very similar to those given by the LC model whereas the histograms for the pensioners show close resemblance to those from the CBD model. Note that after allowing for model error, the final distributions appear unimodal. This is probably due to the fact that the weightings are highly lopsided and that the two mortality models generate very close modes using the CMI data despite the diversity in their model structures. For other models and data sets, if the candidate models produce more different mode values, it might be possible that the simulated final distributions turn out to be multimodal. It would be interesting to apply a variety of other models to different data sets and investigate this possibility in future work.

Some might ask what leads to the result that one model is preferred for one data set but not the other. We take a deeper dive into the data and find that beyond age 87, the death rates of the assured lives, unlike those of the pensioners, did not curve up over age as much as “expected”. Instead, they stayed relatively flat as shown in the age profile of death rates over the selected years for each data set (Figure 3.8). As a result, the CBD model with curvature which relies on a quadratic form to fit the logit death rates could not capture the “discontinuity” sufficiently well. On the other hand, the LC model has a free parameter for modelling the mean death rate for each age and thus performs better. For the pensioners data set, the advantage of the LC model is much reduced. The LC model loses to the CBD model due to the fact that it has more effective parameters and

therefore receives greater penalty in the computation of BIC.

3.5 Risk-neutral Pricing via Maximum Entropy

In this section, we would like to examine whether or how the inclusion of model error would affect the pricing of longevity products. In particular, we consider a relatively simple mortality-linked derivative, namely q -forward (Coughlan et al., 2007). Despite its simple structure, an optimised portfolio of q -forwards has been shown to be an effective hedge for longevity risk (Li and Luo, 2012).

The structure and payoff of q -forward are similar to those of a traditional forward in the financial market. Suppose $q_{x,T}$ is the death probability of a reference population at age x in year T . The corresponding forward price F_x is determined at inception such that the risk-neutral expectation of the difference between $q_{x,T}$ and F_x discounted to time zero at the risk-free rate is equal to zero in the absence of any arbitrage opportunities. Hence,

$$F_x = \mathbf{E}^Q[q_{x,T}] \quad (3.3)$$

To determine the risk-neutral distribution, we will adopt the maximum entropy approach. It was first introduced by Stutzer (1996) into the finance literature and is now gaining traction in the actuarial community. Li (2010) made a cogent argument for the advantages of the maximum entropy approach, or the canonical valuation as he called it, over the other risk-neutral frameworks. First, it does not require subjective decisions on the choice of parameters. Second, it can incorporate as many transaction data on existing securities as are available. Since the life market is still in its early stage, there have been few transactions yet. However, as more deals are made in the future, this approach could readily incorporate the additional market information for valuation. Kogure and Kurachi (2010) and Kogure et

al. (2014) applied this valuation method in a Bayesian setting. For applications of other pricing methods, see, for example, Wang et al. (2008) and Chen et al. (2010).

Now we shall detail the procedures to derive the risk-neutral distribution under the maximum entropy approach. In Section 3, we have generated $n=10,000$ sample paths and for the j^{th} path, we obtain a simulation of the process $\{m_{x,s}^{(j)}\}_{s=1,2,\dots,T}$. The simulations for $j = 1, 2, \dots, n$ could be generated from each model alone or from our proposed integrated approach which accounts for model error.

3.5.1 Estimating the Risk-neutral Distribution

Let π denote the empirical distribution of the n paths simulated. The real-world probability of each path π_j is equal to $1/n$. Our objective is to find π^* , the risk-neutral distribution. According to the maximum entropy principle, π^* should be chosen to minimise the Kullback-Leibler information criterion, that is,

$$\sum_{j=1}^n \pi_j^* \ln \left(\frac{\pi_j^*}{\pi_j} \right) \quad (3.4)$$

subject to the following constraints,

$$\begin{aligned} \pi_j^* &> 0, \forall j \\ \sum_{j=1}^n \pi_j^* &= 1 \\ \sum_{j=1}^n a_x^{(j)} \pi_j^* &= a_x^{market} \end{aligned}$$

The first two constraints ensure that π^* is a proper probability measure. The last constraint says that the risk-neutral measure should give a

market-consistent valuation for the existing securities. The term $a_x^{(j)}$ denotes the fair annuity premium for a person aged x calculated based on the simulated mortality from the j^{th} sample path, whereas a_x^{market} denotes the prevalent market value of the annuity. For demonstration, we consider the same annuity as in Section 3. We assume the market price is £16 for the assured lives and £15 for the pensioners at age 65.

The solution of π^* is given by

$$\pi_j^* = \frac{\pi_j \exp(\gamma a_x^{(j)})}{\sum_{k=1}^n \pi_k \exp(\gamma a_x^{(k)})}, \forall j \quad (3.5)$$

where γ is a constant obtained from numerically solving equation 3.6. It is in fact the Lagrange multiplier for the third (price) constraint in the minimization problem. If more transaction data are available, they can be readily incorporated as additional constraints.

$$a_x^{market} = \frac{a_x^{(j)} \exp(\gamma a_x^{(j)})}{\sum_{k=1}^n \exp(\gamma a_x^{(k)})} \quad (3.6)$$

3.5.2 Results and Analysis

Figure 3.9 shows the forward prices calculated for the assured lives and pensioners at age 65. The contract is taken to commence in 2007 and we have tested all possible maturity dates over the projection period, that is, $T = 1, 2, \dots, 31$. Within each chart, we display three sets of results: (1) the dotted (dot-dashed) line denotes the results from the LC (CBD) model with process error alone; (2) the short (long) dashed line indicates the results from the LC (CBD) model including both parameter error and process error; and (3) the solid line represents the results that take into account model error from our integrated approach.

It can be seen that the pricing curves that account for process error only are always below the corresponding curves that account for both parameter error and process error regardless of the mortality model used. Thus accounting only for process error tends to underestimate the forward price. The difference is more apparent for the CBD model with curvature and for the assured lives data set. As the weights in Section 3.5 would have hinted, the integrated approach leans more toward the LC model for the assured lives but draws closer to the CBD predictions for the pensioners. The divergence between the LC and CBD curves is very wide for the assured lives, showing that the consequence of choosing the “wrong” model would be huge. The divergence reaches the maximum at a maturity of 25 years and beyond that, it stays almost constant. For these two data sets, if one selects a particular model upfront by using the BIC, it seems that one would probably not be much worse off compared to using our integrated approach. However, for other data or models, the integrated approach could give a result that is away from the predictions made by any of the considered models. In that case, it would be dangerous to just choose one model upfront and ignore the existence of model error. Such concerns are not without merit. It should be noted that the CMI data set used here is an industry aggregate. Thus, the extent of model error is not so severe due to the large exposure. For individual insurers, the data sets on hand are of a much smaller coverage and consequently subject to greater model selection issue. Our integrated approach could serve as an effective tool for tackling the problem.

The discussions above, however, are based on one single statistical criterion. When a set of different selection criteria are used, one may expect that there is more competition between candidate models, as a model may excel in one aspect but not the others. In this chapter, we demonstrate a relatively basic application, but the underlying framework is general in

nature and can involve various combinations of different features. Further research is needed for exploring the many other possibilities.

Summary

In short, we have introduced a modification to the semi-parametric bootstrap to account for process error, parameter error, and model error in an integrated manner. We have applied it to the CMI data set for both assured lives and pensioners and used the simulated mortality scenarios to price q -forward contracts via the maximum entropy approach. We find that model selection has a significant impact on risk-neutral valuation and thus it is important to have a proper allowance for model error.

In our study, we have only considered using the BIC as the selection criterion. Future research could test other information criteria with different penalties on the number of parameters. Apart from quantitative requirements, certain qualitative conditions may also be added, such as the biological reasonableness of the projected death rates. One could also explore how having more competing models would affect the results. This would be computationally more expensive but likely to have better error incorporation due to enlarged feasible region. The point estimates would be some “weighted averages” of the competing models and the results would depend on the specific data sets and evaluation criteria. Another interesting line of research is to set a prior probability for each model considered and derive the posterior distributions under a Bayesian framework, as proposed by Cairns (2000, 2014). This approach, however, would require heavy computations and programming. Finally, it would be interesting to apply our integrated method in settings where exposure is relatively small. We would imagine that model selection be more “volatile” and ex-ante selection of a mortality model run a greater risk of misspecification. Application to such data sets would provide further legitimacy to our proposed

method.

In LC and CBD type of models, mortality is expressed in terms of (central) death rates, i.e. $m_{x,t}$ or $q_{x,t}$. Some literature examines other mortality quantities. For instance, Thatcher et al. (1999) focused on the force of mortality $\mu_{x,t}$, the behavior of which especially at advanced ages could be better characterized based on theories consistent with the aging process. De Jong and Marshall (2007) analyzed the Wang-transformed survival probabilities $s_{x,t}$ and found persistent patterns in the distortion parameter that is amenable to forecasting. Certainly, different modelling choices would entail different model strengths and limitations. Further research could look at the tensions between these alternative models and evaluate the impact of model errors accordingly.

FIGURE 3.1: Mortality indices from LC (top left) and CBD (top right and bottom) models using assured lives data.

FIGURE 3.2: Log death rates from LC (top) and CBD (bottom) models using assured lives data.

FIGURE 3.3: Histograms of simulated present values of an annuity from LC (top) and CBD (bottom) models using assured lives data.

FIGURE 3.4: Mortality indices from LC (top left) and CBD (top right and bottom) models using pensioners data.

FIGURE 3.5: Log death rates from LC (top) and CBD (bottom) models using pensioners data.

FIGURE 3.6: Histograms of simulated present values of an annuity from LC (top) and CBD (bottom) models using pensioners data.

FIGURE 3.7: Histograms of simulated present values of an annuity given by the integrated approach using assured lives (top) and pensioners (bottom) data respectively.

FIGURE 3.8: The age profile of death rates at selected years for assured lives (left) and pensioners (right) data.

FIGURE 3.9: Forward prices of View the q -forwards with different maturity dates computed from LC model with process error only (dotted line), LC model with parameter and process error (short dashed line), CBD model with process error only (dot-dashed line), CBD model with parameter and process error (long dashed line) and our integrated approach (solid line) using assured lives (top) and pensioners (bottom) data.

TABLE 3.1: Mean exposures and death rates over 1983–1994 and 1995–2006 at key ages for the assured lives, pensioners and England & Wales population data.

	Assured lives		Pensioners		E&W population	
	1983–1994	1995–2006	1983–1994	1995–2006	1983–1994	1995–2006
Exposures						
65	33,952	24,135	6,863	6,026	505,679	484,449
70	11,320	10,668	16,798	9,599	448,423	439,195
75	6,487	6,866	16,086	8,913	358,272	387,428
80	3,655	3,805	11,420	7,646	261,390	280,849
85	1,461	1,842	4,935	5,295	140,700	172,656
90	463	757	1,286	2,315	52,686	79,091
Death rates						
65	0.01522	0.00954	0.02063	0.01255	0.01993	0.01423
70	0.02501	0.01897	0.03294	0.02077	0.03163	0.02377
75	0.04311	0.03198	0.05704	0.03973	0.04962	0.04021
80	0.07459	0.05611	0.09076	0.06749	0.07989	0.06607
85	0.11928	0.09622	0.14441	0.11189	0.12653	0.11059
90	0.18324	0.14752	0.20706	0.17687	0.20002	0.18306

TABLE 3.2: Proportion of simulated scenarios with a lower BIC value for each model.

	LC	CBD
Assured lives	0.986	0.014
Pensioners	0.089	0.911

Chapter 4

Longevity Risk and Current Developments

Background

Human life expectancy has been increasing in an almost linear fashion with no signs of slowing down (Oeppen and Vaupel, 2002; see Figure 4.1). Significant amount of resources have been poured into research on biology and genetics in an attempt to find cancer cures and longevity genes. With the current state of technology, one needs not to be an optimist to believe that human life would continue to extend. Governments of various countries have already started to combat the challenge of an aging society by, for instance, gradually increasing the retirement age, cutting the benefits under the existing DB pension plans and transitioning from a Pay-As-You-Go (PAYG) pension system to a fully funded one (OECD, 2015).

Of particular interest to actuaries is the risk of longevity, i.e. the uncertainty associated with the increase in human lifespan. This risk threatens not only the financial security of individuals but also the viability of life companies, and private and public pension funds. Since 2006, the market for longevity-linked transactions has been active in the UK. By the end of 2011, £40 billion worth of transactions had been completed out of the estimated potential private sector pension liabilities of £2.1 trillion (Blake et

al., 2013). The other Western countries such as the US, Canada, and Australia are fast catching up. Given the size of the pension sector, both private and public, a huge reservoir of pension liabilities awaits capital market solutions. Table 4.1 provides a snapshot of deals that are above £2 billion in recent years.

Blake et al. (2013) gave a detailed discourse on the burgeoning life market. They went through various types of longevity-linked transactions, such as pension buy-outs, buy-ins, longevity swaps, mortality/longevity bonds, sovereign annuities, and mortality forwards. In particular, they examined the two earlier cases – Swiss Re mortality bond Vita I (2003) and EIB/BNP longevity bond (2004). The former was a success while the latter not well received by the market, the reasons cited being poor product design, unattractive premium, small issue size which precluded liquidity in the secondary market, and pension trustees not receptive to the idea of hedging risks using a bond¹. The market has not since been discouraged. On the contrary, with lessons learnt from previous experiences, new products are continuously rolled out, albeit experimentally, to better meet the needs of the participants, as reflected by more transactions in recent years (Tan et al., 2015; Table 4.1).

Loeys et al. (2007) made a case for the life market. They argued that it has the necessary ingredients to develop into a full-fledged market. In a schematic diagram, they listed the potential market participants. Defined Benefit pension funds and annuity providers are the primary potential longevity buyers.² The primary longevity sellers are life insurance companies who are concerned with declines in mortality improvement. They provide a natural, but imperfect, hedge to the longevity buyers. Besides,

¹For related discussions, see Cowley and Cummins (2005), Blake et al. (2006), Lin and Cox (2008) and Chen and Cummins (2010).

²By “longevity buyers”, Loeys et al. (2007) mean buyers of longevity protections as they have the biggest exposures to longevity risk.

there are life settlement funds, premium finance investors, and pension buy-out funds that are willing to sell longevity protections and take up the risks in exchange for a premium. Another group of participants are the so-called liquidity providers who are attracted by the good return of longevity-linked instruments and their low correlation with conventional capital market assets.

The presence of liquidity providers is indispensable to the making of a successful market. Without them, there would not be enough trading activities to impound information into the prices. The bid-ask spread would be wide and transaction costs high. Some would-be investors might shun away from the market for fear of being unable to sell or buy back at a reasonable price. Thus, one key question is how to attract the liquidity providers, or to put it simply, how to price longevity-linked products so that the return offered is enough to compensate the risks that investors take.

To the general investors from the capital market, the prices they are willing to pay depend much on the correlation (or "marginal benefits") between longevity-linked products and traditional financial assets. If the correlation is truly as low as what is widely believed, adding longevity products to the portfolio would give more diversification benefits and deliver higher Sharpe ratio. As a result, the investors would demand lower compensation for taking on longevity risk. The prevalent view in the academia as well as in the industry is that the correlation between longevity products and conventional assets is fairly low and insignificant (See, for example, Lin and Cox, 2005; Loeys et al., 2007; Ribeiro and Pietro, 2009; Mayhem and Smith, 2011; Tsai et al., 2011).

4.1 Pricing of Longevity Risk

In the literature, most studies employed the risk-neutral pricing method to quantify longevity risk. One of the popular approaches is via the Wang transform. Lin and Cox (2005) first applied this to the pricing of mortality bonds, followed by Denuit et al. (2007). Later works have considered extensions, such as incorporating jumps in the stochastic mortality processes and accounting for parameter uncertainties (Lin & Cox, 2008; Chen & Cox, 2009). Another popular approach is the Esscher-type of transform. Works along this line include Cox et al. (2006), Goovaerts & Laeven (2008) and Chuang & Brockett (2014). There are some alternative risk-neutral pricing methods. Following Milevsky and Promislow (2001) and Dahl (2004), Cairns et al. (2006) introduced a set of market risk parameters for pricing within the Cairns-Blake-Dowd (CBD) model. Kogure and Kurachi (2010) and Kogure et al. (2014) adopted the maximum entropy principle and obtained the risk-neutral distribution with Bayesian MCMC. For these studies, the risk-neutral probability (or transform parameter) is determined by benchmarking against either existing annuities or the earlier marketed longevity products, such as the Swiss Re mortality bond.

Chen and Cummin (2010) took an actuarial pricing approach by loading a risk premium to the expected losses. They used the cubic model to calculate the expected excess return (EER), the part of the premium that compensates the investors for uncertainty over the investment outcome. The EER is calibrated using CAT bond transaction data.

Regardless of the pricing methods, the studies have generally assumed

that longevity risk and financial risk, particularly interest rate risk, are independent. Nonetheless, some authors did express concerns over the validity of the assumption. Cairns et al. (2006) believed that such an assumption is reasonable for "relatively short horizons under normal circumstances" but recognized that "over the very long run, the term structure of interest rates will be influenced by the relative size of capital stock to that of the population, and the latter might be influenced by mortality ... dynamics". Similar view is echoed in Denuit et al. (2007). They discussed possible linkages between demographic changes and stock market returns as postulated by the life-cycle theory. However, they argued that the dependence is indirect. Moreover, the downward pressure on asset prices from the retiring population would be counteracted by capital inflows from regions of young workers.

Remark. It is worth noting that although independence is assumed in the valuation, it might have been indirectly and partially imputed during the calibration stage. For instance, Chen and Cummin (2010) estimated the EER with the CAT bond data, which should contain information on the price of mortality risk as perceived by the capital market investors, possibly including their assessment of the correlation. However, we should be cautious that although both CAT bonds and longevity bonds account for the uncertainties in mortality changes, they are different in nature. The former is mainly concerned with short-term temporal jumps in the mortality of a certain reference group, while the latter focuses on long-term changes in mortality decline (Chen & Cox, 2009). For those that were calibrated against annuities or longevity products, the match of longevity risk is there. However, there could be other margins (or "paddings") that are related to transaction cost, adverse selection, market competition and relative bargaining powers.

In the next two sections, we will review the relevant literature from economics, finance and actuarial science to provide some broad perspectives on the issue. We divide the discussion into two categories, one related to the macroeconomic changes, such as per-capita GDP growth and unemployment rate fluctuations, and the other to the financial market movements, such as equity and bond prices and returns.

4.2 Longevity and Macroeconomic Changes

From a theoretical viewpoint, there are several economic studies which may provide us some insights. De la Croix and Licandro (1999) reasoned that longer life expectancy could have more direct effects on growth, apart from being a proxy for factors that reflect the general productivity of a society, as suggested by Barro and Sala-i-Martin (1995). If the probability of living to old age is higher, it is plausible that individuals would have lower subjective discount rate³, invest more in education, and delay the start of work. At the same time, the rate of human capital depreciation would be lower, prompting further investment in education. If human capital accumulation is an important driver of growth, longer life expectancy would have a positive effect on growth.

More formally, they proposed a continuous-time over-lapping generations (OLG) model⁴ and derived the equilibrium conditions which stipulate that the instantaneous interest rate equates the subjective discount rate and that the optimal schooling time is inversely proportional to the sum of the discount rate and the force of mortality. As such, the effect

³In the recursive preferences setting, a positive longevity shock will lead to lower discount rate and more precautionary savings only under certain conditions (Jang et al., 2016).

⁴In the model, they assumed constant force of mortality, linear utility function, and the existence of perfect annuity markets. The problem of an agent is to choose a consumption series and the length of education so as to optimize the expected utility given the budget constraint, aggregate human capital, wage, and interest rates.

on the growth rate of an increase in life expectancy, characterized by a reduction in the force of mortality, is three-fold: (1) positive through slower depreciation of human capital; (2) positive through greater investment in education that increases per-capita human capital; and (3) negative due to the fact that the work force is older on average. The net effect would be positive for economies with a relatively low life expectancy, but could be negative for more developed economies.

Zhang et al. (2001) focused on mortality improvement at older ages and its implications on the long-run economic growth with and without a social security system. Specifically, they considered a discrete-time OLG model whereby each agent lives for three periods: child, adult and retiree. Survival is uncertain only from middle to old age while fertility is endogenously chosen. Household trades off between the number of children and the welfare of children. They further assumed the existence of a perfect annuity market. The decision problem for a middle-aged agent is to choose an allocation plan for savings, number of children, per-child investment and bequest so as to maximize the expected utility.

Without social security or with fully funded social security, the effect of longevity hinges on how much the parents value the number of children over the welfare of children. If the parents have a stronger preference for the number of children, a rise in life expectancy reduces fertility and promotes human capital investment and per-capita growth, and vice versa. In the case of unfunded social security, a rise in life expectancy raises social security contribution. It also lowers fertility and increases human capital investment and growth for a wide range of plausible parameter values. Given the decreasing trend in fertility worldwide, their findings would suggest a positive relation between longevity and growth.

Turn to the empirical works. Using state-level data, Ruhm (2000 & 2006) found a procyclical pattern of mortality⁵ in US for the 20 years from 1972. That is "recessions are good for [our] health" as he notably put in the title. Miller et al. (2009) reaffirmed the result for a more recent period and analyzed the relationship in greater depth. They found that the effect is less likely due to high stress levels during economic booms than to externalities of business cycles, such as increased traffic activities and pollution. Using the Lee-Carter mortality indexes for 6 OECD countries from 1950 to 2006, Hanewald (2011) also found that a significant portion of the indexes' changes are positively related to per-capita real GDP growth, and negatively to unemployment rate fluctuations. He further did a cointegration test and found evidence of a long-run relationship between the mortality indexes and macroeconomic variables. Given that the economies under study are mostly developed regions, the empirical evidence seems to lend support to the predictions in de la Croix and Licandro (1999).

Sullivan and Wachter (2009), however, found dissimilar results based on quarterly data from the state of Pennsylvania in 1980-2006. They found that displaced workers⁶ tend to experience higher mortality rates even after controlling for possible endogenous layoffs.

4.3 Longevity and Financial Returns

The previous section examines longevity in the wider context of the macroeconomy. In this section, we narrow our focus to look at the relation between longevity and the financial markets. Economists have examined whether demographic characteristics, such as average population age and old-age dependency ratio, have any influence over asset price levels and/or returns. Though such variables are not direct measures of longevity, they

⁵Mortality due to causes like motor accidents, cardiovascular diseases and influenza.

⁶They refer to people who lose their jobs due to plant closings, mass layoffs and other firm-level reductions.

could serve as longevity proxies to say the least.

It has been hypothesized that longevity could affect asset returns through the channel of risk aversion. As an individual becomes older, she would prefer to hold less risky assets such as government bonds than riskier assets like stocks. Bakshi and Chen (1994) found empirical support for this argument. They calibrated the utility function of a representative agent such that the aggregate risk aversion γ_t varies with the average age M_t of the US adult population (above age 19),

$$U(C_t, \gamma_t) = \frac{C_t^{1-\gamma_t}}{1-\gamma_t} \quad (4.1)$$

where C_t is the real consumption of nondurables and services and $\gamma_t = \gamma + \lambda M_t$ is the time-varying risk aversion. Thus, the Euler equation is

$$\mathbf{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_t} (1 + r_t) \right] = 1 \quad (4.2)$$

In doing so, they obtained a better fit for the intertemporal Euler equation for aggregate consumption growth. The coefficient λ was significantly positive regardless of the instruments used. As we know from finance literature that risk aversion is a key determinant of asset returns, the results would suggest that demographic changes could have a substantial impact on asset returns.

In a follow-up regression analysis, they found that the change in average age has predictive power for future stock risk premiums in the presence of dividend yield and consumption growth. Together, these variables could explain 33% of the variations in future risk premium. The change in average age was the most robust among the explanatory variables, providing further evidence for the link between longevity and asset returns.

By considering the more general recursive utility, Jang et al. (2016) studied risky asset demands and consumption in the presence of longevity risk. They found that in the partial insurance case, if the return premium on fixed annuities is high enough, the demand for risky assets would be lower than in the case of no or complete insurance. The effect propagates through the relative risk aversion and elasticity of intertemporal substitution.

Apart from risk aversion, longevity could affect asset returns through the life cycle of investment needs. For instance, based on the post-war US data, Bakshi and Chen (1994) found evidence that as average population ages, real stock (housing) price index increases (decreases), consistent with the hypothesis that older investors demand proportionately more financial assets than real estates.

Bergantino (1998) estimated the age-specific demand for stocks and real estate using data from the Survey of Consumer Finances (SCF). Coupled with the changing demographic structure, he projected the demand over the period from 1966 to 1997 and found that the projected demand is clearly positively related to the level of stock prices.

Poterba (2001), on the other hand, studied the relationship between several demographic variables⁷ and the returns of T-bills, long-term government bonds, and common stocks. He found some correlation between the demographic structure and fixed income asset returns, but the coefficients are sensitive to the time periods. He also tested the demographic variables against asset price levels as measured by price-dividend ratios. His results are similar to those in Bergantino (1998). In this regards, the results are

⁷The variables include median age, average age of those 20+, percentage of population 40-64, ratio of population 40-64 to 65+, ratio of population 40-64 to 20+.

consistent with the idea that longevity risk and financial risk are likely to be related in the long run. Poterba (2004) performed similar econometric tests with an updated data set and the conclusions are largely the same.

More recently, Loeys et al. (2007) and Ribeiro and Pietro (2009) examined the relationship of changes in average mortality rates with equity and bond market returns in US, UK and Japan for various holding periods and found little evidence of correlation. Ribeiro and Pietro (2009) also attempted to back out an implied longevity return series using monthly UK annuity prices conditional on yield curve changes for the period 1991-2009. Still, they found this return metric to be uncorrelated with FTSE returns or 10-year UK Gilt returns.

Exploiting the properties of guarantee annuity, Milevsky (2005) computed an implied longevity yield (ILY) index using weekly Canadian life annuity quotes. The ILY is essentially the internal rate of return over the guarantee period that a person who decides to self-annuitize would have to earn on her portfolio in order to replicate the income from the existing annuity assuming current pricing assumptions remain unchanged. Contrary to Ribeiro and Pietro (2009), Milevsky (2005) found that the index is highly correlated with the long-term government bond returns in Canada. By another different approach, Bauer et al. (2010) found similar results to those in Milevsky (2005). They computed longevity Sharpe ratios as well as risk premiums measured by the Wang transform distortion using UK forward mortality tables, immediate annuity quotes and government bond yields from 1999 to 2006. For the early half period, the correlation between the metrics and stock market returns were as high as -0.8. It reduced somewhat in the second half but still remained a sizable -0.5. As for bond returns, the correlation was insignificant in the early period but became highly significant (0.9) in the later period.

Even results from the more sophisticated methods are mixed. This could be because different methods impose different assumptions in their calculations, because insurance companies inherently have delays in their adjustment of various rates, or because there are some omitted variables that confound the analyses.

Summary

On the whole, the results are not conclusive at best. Under the current stage of the life market development, it is inevitable that we make some assumptions on the (in)dependence between longevity and financial risks in pricing. Otherwise, we may find it almost impossible to arrive at a sensible figure as the transaction price. Nonetheless, it would be beneficial to take a closer look at the validity of the assumption on the relationship between longevity and financial risks. Are they correlated and/or dependent in the first place? If so, by how much and in what form? Linear or non-linear? Over how long a horizon? Maybe we could find some hints from simulations of more complex but highly realistic models as in Hubener et al. (2016). Once we obtain a better understanding, we would be able to perform more precise, even if more sophisticated, valuation of longevity risk. Meanwhile, we could pay attention to the upcoming deals and observe how actual prices change.

FIGURE 4.1: Best practice female life expectancy from 1840 to 2000.

Data are obtained from Oeppen and Vaupel (2002) supplemental table 2.

TABLE 4.1: Large longevity-linked transactions from 2010 to 2015.

Fund / Sponsor	Provider(s)	Solution	Size (bn)	Date
Philips UK Pension Fund	Pension Insurance / Hannover Re	Buyout & longevity reinsurance	GBP 2.4	Nov-15
Scottish & Newcastle PP	Friends Life (with Swiss Re)	Longevity swap & reinsurance	GBP 2.4	Sep-15
Aegon	Canada Life Re	Longevity swap & reinsurance	EUR 6.0	Jul-15
AXA UK Pension Scheme	RGA Re	Longevity swap	GBP 2.8	Jul-15
Delta Lloyd	RGA Re	Index-based longevity swap	EUR 12.0	Jun-15
Bell Canada PP	Sun Life (with SCOR & RGA Re)	Longevity swap & reinsurance	CAD 5.0	Mar-15
Delta Lloyd	RGA Re	Index-based longevity swap	EUR 12.0	Aug-14
BT Pension Scheme	Prudential	Pensioner bespoke longevity swap	GBP 16.0	Jul-14
AkzoNobel	Legal & General / Prudential	Buy-in, bulk annuity & longevity reinsurance	GBP 3.6	Mar-14
Aviva	Swiss Re / Munich Re / SCOR	Pensioner bespoke longevity swap	GBP 5.0	Mar-14
Astra Zeneca	Deutsche Bank / Abbey Life	Pensioner bespoke longevity swap	GBP 2.5	Dec-13
BAE Systems	Legal & General / Hannover Re	Pensioner bespoke longevity swap	GBP 3.2	Feb-13
Verizon	Prudential	Pension liability buy-out	USD 7.0	Oct-12
General Motors	Prudential	Pension liability buy-out	USD 26.0	Jun-12
Aegon	Deutsche Bank	Capital markets longevity swap	EUR 12.0	Feb-12
Rolls Royce	Deutsche Bank / Abbey Life	Pensioner bespoke longevity swap	GBP 3.0	Nov-11
BMW	Deutsche Bank / Abbey Life	Pensioner bespoke longevity swap	GBP 3.0	Feb-10

Data are obtained from www.artemis.bm.

Bibliography

- Andreev, K. F. (2002). *Evolution of the Danish Population from 1835 to 2000*. Vol. Monographs on Population Aging 9. Odense University Press.
- Bakshi, G. and Z. Chen (1994). "Baby boom, population aging, and capital markets". In: *Journal of Business* 67, pp. 165–202.
- Barro, R. J. and X. S. i Martin (1997). "Technological diffusion, convergence, and growth". In: *Journal of Economic Growth* 2.1, pp. 1–26.
- Bauer, D., M. Borger, and J. Rub (2010). "On the pricing of longevity-linked securities". In: *Insurance: Mathematics and Economics* 46, pp. 139–149.
- Bergantino, S. (1998). "Lifecycle investment behavior, demographics, and asset prices". PhD thesis. MIT, Department of Economics.
- Blake, D., A. J. G. Cairns, and K. Dowd (2006). "Living with mortality: longevity bonds and other mortality-linked securities". In: *British Actuarial Journal* 12.1, pp. 153–228.
- Blake, D. et al. (2013). "The new life market". In: *Journal of Risk and Insurance* 80.3, pp. 501–557.
- Booth, H., J. Maindonald, and L. Smith (2002). "Applying Lee-Carter under conditions of variable mortality decline". In: *Population Studies* 56, pp. 325–336.
- Booth, H. et al. (2006). "Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions". In: *Demographic Research* 15, pp. 289–310.
- Bray, I. (2002). "Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 51.2, pp. 151–164.

- Brouhns, N., M. Denuit, and I. V. Keilegom (2005). "Bootstrapping the Poisson log-bilinear model for mortality forecasting". In: *Scandinavian Actuarial Journal* 31, pp. 212–224.
- Brouhns, N., M. Denuit, and J. K. Vermunt (2002a). "A Poisson log-bilinear regression approach to the construction of projected life tables". In: *Insurance: Mathematics and Economics* 31.3, pp. 373–393.
- (2002b). "Measuring the longevity risk in mortality projections". In: *Bulletin of the Swiss Association of Actuaries*, pp. 105–130.
- Cairns, A. J. G. (2000). "A discussion of parameter and model uncertainty in insurance". In: *Insurance: Mathematics and Economics* 27.3, pp. 313–330.
- (2014). "Mortality seminar series: Exploring the future; defining the questions". In: *British Actuarial Journal* 19.3, pp. 650–691.
- Cairns, A. J. G., D. Blake, and K. Dowd (2006). "A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration". In: *Journal of Risk and Insurance* 73, pp. 687–718.
- Cairns, A. J. G. et al. (2009). "A quantitative comparison of stochastic mortality models using data from England and Wales and the United States". In: *North American Actuarial Journal* 13.1, pp. 1–35.
- Cairns, A. J. G. et al. (2011a). "Bayesian stochastic mortality modelling for two populations". In: *ASTIN Bulletin* 41.1, pp. 29–59.
- Cairns, A. J. G. et al. (2011b). "Mortality density forecasts: An analysis of six stochastic mortality models". In: *Insurance: Mathematics and Economics* 48, pp. 355–367.
- Chan, W. S., J. S. H. Li, and J. Li (2014). "The CBD mortality indexes: modeling and applications". In: *North American Actuarial Journal* 18.1, pp. 38–58.

- Chen, H., S. H. Cox, and S. S. Wang (2010). "Is the home equity conversion program in the United States sustainable? Evidence from pricing mortgage insurance premiums and non-recourse provisions using the conditional Esscher transform". In: *Insurance: Mathematics and Economics* 46.2, pp. 371–348.
- Chen, H. and S. H. Cox (2009). "Modeling mortality with jumps: applications to mortality securitization". In: *Journal of Risk and Insurance* 76.3, pp. 727–751.
- Chen, H. and J. D. Cummins (2010). "Longevity bond premiums: The extreme value approach and risk cubic pricing". In: *Insurance: Mathematics and Economics* 46, pp. 150–161.
- Chernick, M. (1999). *Bootstrap methods: A practitioner's guide* (pp. 1-22). New York: Wiley.
- Christensen, K. et al. (2009). "Ageing populations: the challenges ahead". In: *The Lancet* 374.9696, pp. 1196 –1208.
- Chuang, S.-L. and P. L. Brockett (2014). "Modeling and pricing longevity derivatives using stochastic mortality rates and the Esscher transform". In: *North American Actuarial Journal* 18.1, pp. 22–37.
- Coughlan, G. D. et al. (2007). "q-forwards: Derivatives for transferring longevity and mortality risk". In: *JPMorgan Pension Advisory Group*.
- Cowley, A. and J. D. Cummins (2005). "Securitization of life insurance assets and liabilities". In: *Journal of Risk and Insurance* 72.2, pp. 193–226.
- Cox, S. H., Y. Lin, and S. Wang (2006). "Multivariate exponential tilting and pricing implications for mortality securitization". In: *Journal of Risk and Insurance* 73.4, pp. 719–736.
- Currie, I. D. (2006). "Smoothing and forecasting mortality rates with P-splines". In: *Talk given at Institute of Actuaries, June 2006*.
- (2016). "On fitting generalized linear and non-linear models of mortality". In: *Scandinavian Actuarial Journal* 2016.4, pp. 356–383.

- Currie, I. D., M. Durban, and P. H. C. Eiler (2004). "Smoothing and forecasting mortality rates". In: *Statistical Modelling* 4.4, pp. 279–298.
- Czado, C., A. Delwarde, and M. Denuit (2005). "Bayesian Poisson log-bilinear mortality projections". In: *Insurance: Mathematics and Economics* 36, pp. 260–284.
- Dahl, M. (2004). "Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts". In: *Insurance: Mathematics and Economics* 35, pp. 113–136.
- Dahl, M. and T. Møller (2006). "Valuation and hedging of life insurance liabilities with systematic mortality risk". In: *Insurance: Mathematics and Economics* 39, pp. 193–217.
- De Jong, P. and C. Marshall (2007). "Mortality Projection Based on the Wang Transform". In: *ASTIN Bulletin* 37.1, pp. 149–161.
- De la Croix, D. and O. Licandro (1999). "Life expectancy and endogenous growth". In: *Economics Letters* 65, pp. 255–263.
- Delwarde, A., M. Denuit, and P. Eilers (2007). "Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized log-likelihood approach". In: *Statistical Modelling* 7.1, pp. 29–48.
- Denuit, M. (2009). *Dynamic life tables: construction and applications*. Munich, Germany.
- Denuit, M., P. Devolder, and A.-C. Goderniaux (2007). "Securitization of longevity risk: pricing survivor bonds with wang transform in the lee-carter framework". In: *Journal of Risk and Insurance* 74.1, pp. 87–113.
- Dowd, K. et al. (2010). "Backtesting stochastic mortality models: An ex post evaluation of multiperiod ahead density forecast". In: *North American Actuarial Journal* 14.3, pp. 281–298.
- Dowd, K. et al. (2011). "A gravity model of mortality rates for two related populations". In: *North American Actuarial Journal* 15.2, pp. 334–356.
- Fries, J. F. (1980). "Aging, natural death and the compression of morbidity". In: *New England Journal of Medicine* 303.3, pp. 130–135.

- Goovaerts, M. J. and R. J. Laeven (2008). "Actuarial risk measures for financial derivative pricing". In: *Insurance: Mathematics and Economics* 42, pp. 540–547.
- Haberman, S. and A. Renshaw (2011). "A comparative study of parametric mortality projection models". In: *Insurance: Mathematics and Economics* 48.1, pp. 35–55.
- Hanewald, K. (2011). "Explaining mortality dynamics". In: *North American Actuarial Journal* 15.2, pp. 290–314.
- Hayflick, L. (1996). *How and why we age*. New York: Ballantine Books.
- (2000). "The future of ageing". In: *Nature* 408, pp. 267–269.
- Hubener, A., R. Maurer, and O. S. Mitchell (2016). "How family status and social security claiming options shape optimal life cycle portfolios?" In: *Review of Financial Studies* 29.4, pp. 937–978.
- Hyndman, R. J., H. Booth, and F. Yasmeen (2013). "Coherent mortality forecasting: the product-ratio method with functional time series models". In: *Demography* 50.1, pp. 261–283.
- Jang, B.-G., H. K. Koo, and Y. Rhee (2016). "Asset demands and consumption with longevity risk". In: *Economic Theory* 62.3, pp. 587–633.
- Jarner, S. F. and E. M. Kryger (2011). "Modelling adult mortality in small populations: the SAINT model". In: *ASTIN Bulletin* 41.2, pp. 377–318.
- Jarner, S. F., E. M. Kryger, and C. Dengsoe (2008). "The evolution of death rates and life expectancy in Denmark". In: *Scandinavian Actuarial Journal* 108, pp. 147–173.
- Kogure, A., K. Kitsukawa, and Y. Kurachi (2009). "A Bayesian comparison of models for changing mortality towards evaluating the longevity risk in Japan". In: *Asia Pacific Journal of Risk and Insurance* 3, pp. 1–22.
- Kogure, A. and Y. Kurachi (2010). "A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions". In: *Insurance: Mathematics and Economics* 46, pp. 162–172.

- Kogure, A., J. Li, and S. Kamiya (2014). "A Bayesian multivariate risk-neutral method for pricing reverse mortgages". In: *North American Actuarial Journal* 18.1, pp. 242–257.
- Koissi, M. C. and A. F. Shapiro (2006). "Fuzzy formulation of the Lee-Carter model for mortality forecasting". In: *Insurance: Mathematics and Economics* 39, pp. 287–309.
- Koissi, M. C., A. F. Shapiro, and G. Hognas (2006). "Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval". In: *Insurance: Mathematics and Economics* 38.1, pp. 1–20.
- Kullback, S. and R. A. Leibler (1951). "On information and sufficiency". In: *Annals of Mathematical Statistics* 22, pp. 79–86.
- Lee, R. D. and L. R. Carter (1992). "Modeling and forecasting US mortality". In: *Journal of the American Statistical Association* 87.419, pp. 659–671.
- Lee, R. D. and T. Miller (2001). "Evaluating the performance of the Lee-Carter method for forecasting mortality". In: *Demography* 38.4, pp. 537–549.
- Li, J. (2010a). "Projections of New Zealand mortality using the Lee-Carter model and its augmented common factor extension". In: *New Zealand Population Review* 36, pp. 27–53.
- (2013). "A Poisson common factor model for projecting mortality and life expectancy jointly for females and males". In: *Population Studies* 67.1, pp. 111–126.
- (2014a). "A quantitative comparison of simulation strategies for mortality projection". In: *Annals of Actuarial Science* 8, pp. 281–297.
- (2014b). "An application of MCMC simulation in mortality projection for populations with limited data". In: *Demographic Research* 30, pp. 1–48.

- Li, J. S. H. (2010b). "Pricing longevity risk with the parametric bootstrap: A maximum entropy approach". In: *Insurance: Mathematics and Economics* 47, pp. 176–186.
- Li, J. S. H., W. S. Chan, and S. H. Cheung (2011). "Structural changes in the Lee-Carter mortality indexes". In: *North American Actuarial Journal* 15.1, pp. 13–31.
- Li, J. S. H. and A. Luo (2012). "Key q-duration: a framework for hedging longevity risk". In: *ASTIN Bulletin* 42.2, pp. 413–452.
- Li, N. and R. Lee (2005). "Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method". In: *Demography* 42.3, pp. 575–594.
- Li, N., R. Lee, and S. Tuljapurkar (2004). "Using the Lee-Carter method to forecast mortality for populations with limited data". In: *International Statistical Review* 72.1, pp. 19–36.
- Lin, Y. and S. H. Cox (2005). "Securitization of mortality risks In life annuities". In: *Journal of Risk and Insurance* 72.2, pp. 227–252.
- (2008). "Securitization of catastrophe mortality risks". In: *Insurance: Mathematics and Economics* 42.4, pp. 628–637.
- Loeys, J., N. Panigirtzoglou, and R. M. Ribeiro (2007). *Longevity: a market in the making*. Tech. rep. JP Morgan.
- MacMinn, R. (2003). *International mortality comparisons*. Presentation to the Society of Actuaries. Vancouver. URL: www.journalofriskandinsurance.org.
- Manton, K. G., L. Corder, and E. Stallard (1997). "Chronic disability trends in elderly United States populations: 1982-1994". In: *Proceedings of the national Academy of Sciences* 94.6, pp. 2593–2598.
- Mayhew, L. and D. Smith (2011). "Human survival at older ages and the implications for longevity bond pricing". In: *North American Actuarial Journal* 15.2, pp. 248–265.

- McCullagh, P. and J. A. Nelder (1989). *Generalized linear models*. 2nd ed. London: Chapman and Hall.
- McNown, R. and A. Rogers (1989). "Forecasting mortality: A parameterized time series approach". In: *Demography* 26.4, pp. 645–660.
- Milevsky, M. A., S. D. Promislow, and V. R. Young (2006). "Killing the law of large numbers: mortality risk premiums and the sharpe ratio". In: *Journal of Risk and Insurance* 73.4, pp. 673–686.
- Milevsky, M. A. (2005). "The implied longevity yield: a note on developing an index for life annuities". In: *Journal of Risk and Insurance* 72.2, pp. 301–320.
- Miller, D. L. et al. (2009). "Why are recessions good for your health?" In: *American Economic Review* 99.2, pp. 122–127.
- Oeppen, J. and J. W. Vaupel (2002). "Broken Limits to Life Expectancy". In: *Science* 296.5570, pp. 1029–1031.
- Plat, R. (2009). "Stochastic portfolio specific mortality and the quantification of mortality basis risk". In: *Insurance: Mathematics and Economics* 45, pp. 123–132.
- Poterba, J. M. (2001). "Demographic structure and asset returns". In: *The Review of Economics and Statistics* 83.4, pp. 565–584.
- (2004). "The impact of population aging on financial markets". In: *NBER working paper*.
- QIS5 Technical Specifications (2010). European Insurance and Occupational Pension Authority.
- Renshaw, A. and S. Haberman (2003). "Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 52.1, pp. 119–137.
- Renshaw, A. E. and S. Haberman (2006). "A cohort-based extension to the Lee-Carter model for mortality reduction factors". In: *Insurance: Mathematics and Economics* 38, pp. 556–570.

- (2008). “On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling”. In: *Insurance: Mathematics and Economics* 42, pp. 797–816.
- (2011). “A comparative study of parametric mortality projection models”. In: *Insurance: Mathematics and Economics* 48.1, pp. 35–55.
- Ribeiro, R. and V. di Pietro (2009). *Longevity risk and portfolio allocation*. Tech. rep. JP Morgan.
- Richards, S. J., J. G. Kirkby, and I. D. Currie (2006). “The importance of year of birth in two-dimensional mortality data”. In: *British Actuarial Journal* 12, pp. 5–38.
- Ricklefs, R. E. and C. E. Finch (1995). *Aging: A natural history*. Scientific American Library.
- Ruhm, C. J. (2000). “Are recessions good for your health?” In: *Quarterly Journal of Economics* 115, pp. 617–650.
- (2007). “A healthy economy can break your heart”. In: *Demography* 44, pp. 829–848.
- Stutzer, M. (1996). “A simple nonparametric approach to derivative security valuation”. In: *Journal of Finance* 51, pp. 1633–1652.
- Tan, C. I. et al. (2014). “Parametric mortality indexes: From index construction to hedging strategies”. In: *Insurance: Mathematics and Economics* 59, pp. 285–299.
- Tan, K. S., D. Blake, and R. MacMinnl (2015). “Longevity risk and capital markets: the 2013–14 update”. In: *Insurance: Mathematics and Economics* 63, pp. 1–11.
- Thatcher, A. R. (1999). “The long-term pattern of adult mortality and the highest attained age”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 162.1, pp. 5–43.
- Tsai, J. T., L. Y. Tzeng, and J. L. Wang (2011). “Hedging longevity risk when interest rates are uncertain”. In: *North American Actuarial Journal* 15.2, pp. 201–211.

- Wang, L., E. A. Valdez, and J. Piggott (2008). "Securitization of longevity risk in reverse mortgages". In: *North American Actuarial Journal* 12.4, pp. 345–371.
- Willets, R. C. (2004). "The cohort effect: insights and explanations". In: *British Actuarial Journal* 10, pp. 833–877.

Appendix A

Estimation Procedure

The iterative updating procedures are designed to minimise the deviance function.

$$deviance = \sum_{x,t,i} 2 \left[d_{x,t,i} \ln \left(\frac{d_{x,t,i}}{\hat{d}_{x,t,i}} \right) - d_{x,t,i} + \hat{d}_{x,t,i} \right] \quad (A.1)$$

where $d_{x,t,i}$ is the observed number of deaths at age x in year t for gender i and $\hat{d}_{x,t,i}$ is the corresponding fitted number of deaths. Or equivalently one can maximise the log-likelihood function

$$l = \ln L = \sum_{x,t,i} d_{x,t,i} \ln(\hat{d}_{x,t,i}) - \hat{d}_{x,t,i} - \ln(d_{x,t,i}!) \quad (A.2)$$

In general, the parameters are updated by the equation $\theta^* = \theta - \frac{\partial l / \partial \theta}{\partial^2 l / \partial \theta^2}$ (the Newton-Raphson method). We illustrate below the procedures for model VII as an example. They can be tweaked easily to suit the other models.

1. Obtain the estimates of $a_{x,i}$, B_x , K_t , $b_{x,i,j}$, and $k_{t,i,j}$ according to the steps described in Li (2013) and treat them as given.
2. Initialise the cohort parameters with $\hat{g}_h = 0$, $\hat{c}_x = 1/90$ and the weighting matrix $w_{x,t} = 0$ if $(t - x)$ belongs to the first and last five cohorts and 1 otherwise. Update $\hat{d}_{x,t,i}$.
3. Update $\hat{g}_h^*$ and normalize with $\sum_h \hat{g}_h^* = 0$. Update $\hat{d}_{x,t,i}$.

4. Update $\hat{c}_x^*$ and normalize with $\sum_x \hat{c}_x^* = 1$. Update $\hat{d}_{x,t,i}$.
5. Compute the deviance or log-likelihood function.
6. Repeat (3) to (5) until the deviance or log-likelihood function converges.

Appendix B

Other Residual Plots and Parameter Estimates

FIGURE B.1: Standardised residuals against age for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.

FIGURE B.2: Standardised residuals against age for PLCM (left), PCFM (middle) and selected PCFC (right), France.

FIGURE B.3: Standardised residuals against calendar year for PLCM (left), PCFM (middle) and selected PCFC (right), Australia.

FIGURE B.4: Standardised residuals against calendar year for PLCM (left), PCFM (middle) and selected PCFC (right), France.

FIGURE B.5: Plots of the PCFC Model III parameter estimates, Australia.

FIGURE B.6: Plots of the PCFC Model III parameter estimates, France.