Modeling and Managing Longevity Risk: Models and Applications

by

Yanxin Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

With the threat of longevity risk to the insurance industry becoming increasingly apparent in recent years, insurers and reinsurers are concerned about how to better model and manage longevity risk. However, modeling and managing longevity risk is not trivial, due in part to its systematic nature and in part to the excessive amount of risk factors that constitute the risk. The theme of this thesis is modeling and managing longevity risk. In particular, this thesis focuses on four types of uncertainties among all possible risk factors. These four risk factors include 1) mortality jump risk; 2) longevity drift risk; 3) population basis risk; and 4) cohort mismatch risk.

In the current literature, a number of stochastic mortality models with transitory jump effects have been proposed to capture mortality jump risk. Rather than modeling the age pattern of jump effects explicitly, most of the existing models assume that the distributions of jump effects and general mortality improvements across ages are identical. Nevertheless, this assumption does not appear to be in line with what can be observed from historical data. In this thesis, we addressed this disconnect by introducing a Lee-Carter variant that captures the age pattern of mortality jumps by a distinct collection of parameters. The model variant was then further generalized to permit the age pattern of jump effects to vary randomly. We illustrated the two proposed models with mortality data from the United States and English and Welsh populations, and further used these data to value hypothetical mortality bonds with similar specifications to the Atlas IX Capital Class B note that was launched in 2013. The features we considered were found to have a significant impact on the estimated prices.

We then explored longevity drift risk, which is the uncertainty about the mortality trend itself. We tackled longevity drift risk by introducing the locally-linear CBD model in which the drifts that govern the expected mortality trend are allowed to follow a stochastic process. Compared to the original CBD model, this specification results in median forecasts that are more consistent with recent trends and more robust relative to changes in the data sample period. Furthermore, the proposed model also yields wider prediction intervals that may better reflect the possibilities of future trend changes. To mitigate the risk associated with changes in drifts, we proposed a new hedging method called the generalized state-space hedging method which demands less stringent assumptions. The proposed method allows hedgers to extract more hedge effectiveness out of a hedging instrument, and is therefore useful when there are only a few traded longevity securities in the market.

To incorporate population basis risk, we further extended the proposed generalized state-space hedging method to a multi-population setting. In this extended hedging method, the hedging strategy is derived by first reformulating the assumed multi-population

stochastic mortality model in a state-space representation, and then considering the sensitivities of the hedge portfolio and the liability being hedged to all relevant hidden states. Inter alia, this method allowed us to decompose the underlying longevity risk into components arising solely from the hidden states that are shared by all populations and components stemming exclusively from the hidden states that are population-specific. The latter components collectively represent an explicit measure of the population basis risk involved. Through this measure, a new metric called standardized basis risk profile was developed. This metric allowed us to assess the relative levels of population basis risk that q-forwards with different reference populations, reference ages, and times-to-maturity may lead to. The proposed methodologies were illustrated using real mortality data from various national populations.

Similar to population basis risk, cohort mismatch risk is another risk that is related to population differences when conducting an index-based longevity hedge. It arises when the hedger chooses to link hedging instruments to different cohorts. Although existing index-based longevity hedging strategies mitigate the risk associated with period effects, they often overlook the risk associated with different cohorts. The negligence of cohort effects may lead to sub-optimal hedge effectiveness if the liability being hedged is a deferred pension or annuity which involves cohorts that are not covered by the data sample. We proposed a new hedging strategy that incorporates both period and cohort effects. The resulting longevity hedge is a value hedge which reduces the uncertainty surrounding the τ-year ahead value of the liability being hedged in terms variance or Value-at-Risk. We further developed a method to expedite the evaluation of a value longevity hedge. By utilizing the fact that the innovations of the stochastic processes for the period and cohort effects are not serially correlated, the proposed method avoids the need for nested simulations that are generally required when evaluating a value hedge.

Acknowledgements

I would like to take this opportunity to express the deepest gratitude to my supervisor and dear friend, Johnny Siu-Hang Li, for his support and guidance over the years. I am also extremely grateful to Wai-Sum Chan and Rui Zhou for their helpful suggestions on my professional development.

Special thanks to my thesis committee members - Fang Yang, Chengguo Weng, Pascal Poupart and Hailiang Yang - for their valuable participation and suggestions, to all my friends, for their help and encouragement during my PhD studies.

Finally, I would like to thank my family, for their love, accompany, and the trust in me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Over the past century, human mortality has undergone substantial improvement. In short, people are living longer and longer. For developed countries such as Canada and Japan, life expectancy has increased dramatically. For example, according to the data provided by the Human Mortality Database (2015), the life expectancy of Canadian male and female offspring was approximately 62 and 66 respectively in 1940, while in 2011, newborn males and females in Canada could expect to live to 80 and 84 respectively. According to a report from the National Institute on Aging[1], a steady increase in global life expectancy has been observed since World War II. A major transition in human health is taking place around the world at different rates and along different pathways.

Although the worldwide dramatic increase in life expectancy over the 20th century can be regarded as one of the greatest achievements of human society, the uncertainty associated with the increase in life expectancy could largely affect the financial strength of the insurance industry. Let us take defined-benefit pension plan as an example. The uncertainty associated with the increase in life expectancy has a significant effect on the pension plans, as the longer individuals live, the larger the pension liabilities will be. Typically, we use longevity risk to refer to the adverse financial consequences that arise when individuals live longer than expected. The threat of longevity risk to the insurance industry has become more apparent in recent years, due in part to the current low-yield environment following the financial crisis of 2007-08 along with the more conservative mortality improvement scales that have been recently introduced by the actuarial profession (Canadian

[1]Available at: https://www.nia.nih.gov/research/publication/global-health-and-aging/living-longer

Institute of Actuaries, 2014; Continuous Mortality Investigation Bureau, 2009a,b; Society of Actuaries, 2014).

By definition, longevity risk contains two important aspects: the uncertainty underlying human mortality and the adverse financial consequences for insurance companies and pension plan providers. Therefore, when studying longevity risk, we need to focus on both aspects. On the one hand, a stochastic mortality model that can accurately measure the underlying uncertainty is essential, as that, from a statistical viewpoint, the uncertainty underlying the mortality rate is the cause of longevity risk. On the other hand, for risk management purposes we also need to investigate how we can efficiently manage adverse financial consequences.

On the modeling front, a number of stochastic mortality models have been proposed to quantify the uncertainty related to the mortality rate and provide mortality forecasts. The most popular single-population models include the Lee-Carter model introduced by Lee and Carter (1992), the Cairns-Blake-Dowd model introduced by Cairns et al. (2006), and the collection of models (Models M1-M8) considered by Cairns et al. (2009,2011a) and Dowd et al. (2010a,b). Some of those models have been extended to model multiple populations, including the augmented common factor model by Li and Lee (2005), the two-population Cairns-Blake-Dowd model by Cairns et al. (2011b), and the gravity model by Dowd et al. (2011a). Many other mortality models have also been proposed to incorporate a certain risk, such as the mortality jump model by Chen and Cox (2009) which integrates a jump process into the Lee-Carter model.

On the management front, solutions for hedging longevity risk can be divided into two categories: customized hedge and index-based hedge. A customized longevity hedge is based on the actual mortality experience of the individuals associated with the liability being hedged; as such, it can eliminate all longevity risk. However, its disadvantages of being more costly and lacking liquidity and transparency have made a customized longevity hedge less attractive to investors in capital markets. Different from customized hedge, an indexed-based longevity hedge is based on a broad-based mortality index which reflects the actual mortality experience of a larger pool of individuals such as a national population. The index-based longevity hedge emerges from the increasing demand for longevity risk transfers and is more attractive to investors in capital markets for its liquidity and transparency.

The majority of the longevity risk transfers executed to date are customized longevity transactions, such as bespoke longevity swaps within the insurance industry. However, as the insurance industry cannot take unlimited amount of risk, the market for index-based mortality derivatives has started to grow, thereby allowing longevity risk to be transferred from insurance market to capital market. The derivation of strategies for optimizing an

index-based longevity hedge is generally based on sensitivity matching. In particular, hedging strategies have been derived by matching the sensitivities of the liability being hedged and the portfolio of hedging instruments with respect to changes in the underlying mortality rates. This method is similar to the delta-hedging method in financial literature.

The task of modeling and managing longevity risk is challenging. First, the nature of longevity risk is systematic and affects all policies. We cannot simply apply the law of large numbers to eliminate longevity risk. Second, there are too many types of uncertainty that constitute longevity risk. Types of uncertainty related to the mortality model include mortality jump risk (related to mortality jumps), longevity diffusion risk (related to the uncertainty surrounding the mortality trend), longevity drift risk (related to the uncertainty of the mortality trend), model risk, and parameter risk. Types of uncertainty related to the hedging of longevity risk include population basis risk, cohort mismatch risk, Poisson risk (also known as small sampling risk), and recalibration risk.

The theme of this thesis is modeling and managing longevity risk. In particular, this thesis focuses on the following four types of uncertainty among all possible risk factors: 1) mortality jump risk, 2) longevity drift risk, 3) population basis risk, and 4) cohort mismatch risk. The former two factors arise from the inadequacy of current models, while the latter two factors arise as consequences for risk management purposes.

Mortality Jump Risk

The first source of uncertainty we consider in this thesis is mortality jump risk. The dynamic of human mortality over time has been subject to short-term mortality jumps. Typically, we use mortality jump to describe the phenomenon of the level of mortality rate over a certain period changing dramatically in relation to the neighboring years. These mortality jumps may be caused by some catastrophic events such as wars (World War I and World War II) and influenza pandemics (1918 Spanish Flu and 1957-58 Asian Flu).

In general, catastrophic mortality events have three features. First, the presence of these events is infrequent, with only one or two catastrophic events observed over a long period. Second, the impact of these mortality events is catastrophic. The occurrence of these events could significantly affect the level of human mortality and trigger a large amount of death claims. For example, the Spanish Flu in 1918 infected 50% of the world's population and caused the death of 40-50 million people (Crosby 1976). Third, the impacts of these catastrophic events fade out very quickly with the mortality level usually recovering to the normal level after only a few years.

The occurrence of these catastrophic mortality events could cause a large amount of deaths which could trigger a large number of unexpected death claims and thereby threaten

the financial strength of the life insurance industry. In recent years, catastrophic mortality bond is often used by insurers and reinsurers as a risk mitigation tool that can help cede exposures to extreme mortality risk. The first of such bonds was called Vita I and was issued by Swiss Re in 2003 to reduce exposure to a catastrophic mortality deterioration in five populations. This bond was regarded as a huge success and led to the issuance of many other catastrophic mortality bonds (see Blake et al., 2006b, 2013).

Numerous mortality models have been proposed to incorporate certain features of mortality jumps. The majority of the mortality jump models aim to provide a better fit to certain features of jump effect including frequency, severity, and correlation across different populations.

One feature that is important but often left unmodeled is the age pattern of jump effect, or how the effect of a mortality jump is distributed among different ages. Most existing models assume that the age pattern of mortality jumps is identical to that of general mortality improvements. Although this assumption eases the difficulty in estimating the model, it is counter-intuitive because mortality jumps can be caused by various events (such as pandemics) which would each affect different ages differently. Therefore, a distinct collection of parameters is required to characterize the age pattern of mortality jumps. In addition, the age pattern of mortality jumps also affects the pricing of catastrophic mortality bonds. In a catastrophic mortality bond, the principal repayment is not guaranteed as it depends on a pre-defined mortality index. The pre-defined mortality index is usually calculated as a weighted average of mortality rates from different age groups. If the age pattern of mortality jumps is calibrated inaccurately, then it could largely affect the pricing of catastrophic mortality bonds.

Longevity Drift Risk

The second source of uncertainty considered in this thesis is longevity drift risk. While Human mortality improves overtime, what is the rate of improvement? Longevity drift risk is the risk associated with the trend in mortality improvement. Compared to mortality jump risk, longevity drift risk affects mortality dynamics in the long run. It can be simply understood as the uncertainty related to the speed of mortality improvement and as an important aspect of uncertainty masked within longevity risk.

Longevity drift risk emerges from trend changes in mortality. Profound evidence for trend changes in mortality has been found all over the world (Gallop, 2006; Kannisto et al., 1994; Vaupel, 1997). Furthermore, the existence of trend changes has been tested by numerous statistical tests (Li et al., 2011; Ahmadi and Li, 2014; O'Hare and Li, 2014). However, in the current literature trend changes is not generally regarded as a risk. Most

of the models that incorporate trend changes reflect historical trend changes, but they do not allow mortality trend to change in the future in a random manner. Therefore, the stochastic nature of the drift may be a desirable property for models that capture drift risk.

Managing longevity drift risk is a new concept first introduced by Cairns (2013). In his work, he extends the traditional delta-hedging to delta-nuga hedging to additionally reduce the exposure to drift risk. However, there are several constraints when applying the delta-nuga hedging method. First, the delta-nuga hedging method is sub-optimal if the linearity assumption does not hold; second, the number of hedging instruments is restricted; third, the delta-nuga hedging method is subject to the singularity problem[2]. A hedging method that can mitigate these limitations has yet to be found.

Population Basis Risk

The third source of uncertainty considered in this thesis is population basis risk, which is a risk that is related to population differences. Population basis risk arises from an index-based hedge when the hedging population differs from the reference population of the hedging instrument.

Population basis risk is inevitable in an index-based longevity hedge. Although a large number of multi-population models have been proposed to capture the dependence within different populations, only a few studies attempt to measure population basis risk. According to current literature, the methods of measuring population basis risk are limited. Previous studies on measuring population basis risk typically follow the framework set out by Coughlan et al. (2011). In this framework, population basis risk is measured by comparing the resulting hedge effectiveness between two situations: the absence or presence of basis risk. This method of measuring basis risk, which can be regarded as a post-simulation approach, is heavily reliant on simulations. It is clear that an analytical method for analysing population basis risk is missing; thus, we have no way of knowing what constitutes basis risk.

Some researchers have derived strategies that incorporate population basis risk (Dowd et al., 2011a; Li and Hardy, 2011; Li and Luo, 2012; Zhou and Li, 2014). However, the previous methods are all restricted to specific models. A more general framework for hedging population basis risk has yet to be discovered.

[2]These three limitations are discussed in more detail in Chapter 3

Cohort Mismatch Risk

The final source of uncertainty considered in this thesis is cohort mismatch risk. Cohort effects refer to the observed phenomena that individuals born in particular generations have experienced more rapid mortality improvement than their adjacent generations. Such effects have been found significant in many countries including the United States and England and Wales. Cohort effects are also known as year-of-birth effects because individuals born in the same year experience the same cohort effects, while individuals born in different years experience different cohort effects.

Similar to population basis risk, cohort mismatch risk is another risk related to population differences when conducting an index-based longevity hedge. Cohort mismatch risk arises when the hedger chooses to link hedging instruments to different cohorts. Existing index-based longevity hedging strategies mitigate the risk associated with period (time-related) effects, but often overlook cohort effects. Only a few studies (Li and Luo, 2011; Cairns et al., 2014; Cairns, 2013) have considered cohort effects when deriving hedging strategies. However, no studies have indicated how cohort mismatch risk can be handled, as cohort effects are either fixed or incorporated indirectly.

As the market for index-based mortality derivatives is still quite far from being large and liquid, the number of tradable longevity products is limited. Commonly the hedging population does not come from the same cohort as the hedging instruments. A hedging method that can eliminate the uncertainty associated with cohort effects would be very useful in the current stage of the longevity market.

1.2 Objectives and Outline of the Thesis

This thesis explores four of the many risk factors that constitute longevity risk. The four types of risk being considered are mortality jump risk, longevity drift risk, population basis risk, and cohort mismatch risk discussed in Chapter 2 to Chapter 5 respectively. Each chapter encompasses two parts of modeling and application.

In Chapter 2, we explore mortality jump risk. In particular, we focus on the age pattern of mortality jump effect. We study how the jump effect is distributed among ages and how the features of different age patterns affect the pricing of catastrophic mortality bonds. Two model variants of the Lee-Carter model are proposed to capture the variations in the age patterns of catastrophic mortality jumps. An innovative "Route II" approach is applied to estimate the mortality jump models. We illustrate the two proposed models with mortality data from the United States and English and Welsh populations. We then

apply the new model variants to pricing mortality-linked bonds and identify the problems that may arise if a constant age pattern is assumed.

In Chapter 3, we explore longevity drift risk. We fist investigate empirical and statistical evidence for stochastic drifts. We then address longevity drift risk by proposing a locally-linear mortality model in which the drifts that govern the expected mortality trend are allowed to be stochastic. We study the forecasting performance and the robustness of the proposed model. A new hedging method, the generalized state-space hedging method, is developed to immunize a portfolio against drift risk. A hypothetical example is provided to illustrate the proposed hedging method.

In Chapter 4, we explore population basis risk. We further extend the generalized state-space hedging method introduced in Chapter 3 to a multi-population setting. Using the proposed hedging method, we analytically decompose the portfolio variance and study the relationship between the hedge effectiveness and the composition of hedging strategies. A new metric called standardized basis risk profile is developed. This metric allows us to assess the relative levels of population basis risk that q-forwards with different reference populations, reference ages, and times-to-maturity may lead to. The proposed methodologies are illustrated using real mortality data from various national populations.

In Chapter 5, we explore cohort mismatch risk. We propose a new hedging strategy that incorporates both period and cohort effects. Using the proposed method, one can create a value hedge for a deferred annuity liability which involves cohort effects that are not yet realized as of the time when the hedge is established. The risk measures we consider include variance and Value-at-Risk. We further develop a method to expedite the evaluation of a value longevity hedge. By utilizing the fact that the innovations of the stochastic processes for the period and cohort effects are not serially correlated, the proposed method avoids the need for nested simulations that are generally required when evaluating a value hedge. Finally, we present the baseline empirical results and perform several sensitivity tests.

In Chapter 6, we conclude the thesis with some suggestions for further research.

Chapter 2

The Age Pattern of Transitory Mortality Jumps and Its Impact on the Pricing of Catastrophic Mortality Bonds

2.1 Introduction

The dynamics of human mortality over time are subject to short-term jumps. These jumps may be caused by influenza pandemics, most notably the Spanish flu in 1918-20 that is estimated to have infected 50% of the world's population and led to a total mortality of 40-50 million (Crosby 1976). More recently, the Asian flu in 1957-8 is believed to have killed approximately 1 million persons in total (Dauer and Serfling, 1961; Pyle, 1986; Potter, 2001). It is reasonable to assume that similar influenza pandemics will occur in future, because there is an unlimited reservoir of influenza subtypes. Also, for reasons such as interspecies transmission, intraspecies variation and altered virulence, the timings and severities of future pandemics (and hence mortality jumps) are unpredictable (Cox et al., 2003; Webster et al., 1997).

Mortality jumps are infrequent, but their occurrence could trigger a large number of unexpected death claims, thereby affecting the financial strength of the life insurance industry. Stracke and Heinen (2006) estimated that the worst pandemic would result in approximately €45 billion of additional claims expenses in Germany. This amount is equivalent to five times the total annual gross profit or 100 percent of the policyholder

bonus reserves in the German life insurance market. Toole (2007) found that in a severe pandemic scenario, additional claims expenses would consume 25 percent of the risk based capital (RBC) of the entire U.S. life insurance industry. This finding means that companies with less than 100 percent of RBC would be at an increased risk of insolvency. In recent years, a number of reinsurers have used catastrophic mortality bonds as a risk mitigation tool. The first of such bonds was called Vita I, issued by Swiss Re in 2003 to reduce its exposure to a catastrophic mortality deterioration in five populations. With a full subscription, this bond was regarded as a huge success and led to many other catastrophic mortality bonds being issued (see Blake et al., 2006b, 2013).

To model extreme mortality risk and value catastrophic mortality bonds, researchers have developed a number of stochastic mortality models that incorporate jump effects. These models include the contributions by Bauer and Kramer (2009), Biffis (2005), Chen (2013), Chen and Cummins (2010), Chen and Cox (2009), Chen et al. (2010, 2013a), Cox et al. (2006, 2010), Deng et al. (2012), Hainaut and Devolder (2008), Lin and Cox (2008), Lin et al. (2013) and Zhou et al. (2013a). Several features of mortality jumps have been studied in great depth. In terms of jump occurrence, Chen and Cox (2009) and Chen et al. (2010) used independent Bernoulli distributions, Cox et al. (2006) considered Poisson jump counts, whereas Lin and Cox (2008) utilized a discrete-time Markov chain. In terms of jump severity, Chen and Cox (2009) and Chen et al. (2010) made use of normal distributions, Chen and Cummins (2010) applied the extreme value theory, while Chen (2013) and Deng et al. (2012) considered double-exponential jumps. In terms of correlations across different populations, Chen et al. (2013a) used a factor-copula method, Lin et al. (2013) built a model with correlated Brownian motions, whereas Zhou et al. (2013a) considered a multinomial approach.

One feature that has not been studied extensively is the age pattern of mortality jumps, that is, how the effect of a mortality jump is distributed among different ages. Most of the existing models are either constructed for modeling aggregate mortality indexes that are based on total annual death and exposure counts, or configured in such a way that the age pattern of mortality jumps is identical to that of general mortality improvements. To discern the potential limitations of these modeling approaches, let us perform an exploratory analysis on some of the short-term mortality jumps that occurred in the U.S. and England and Wales since 1901. We first apply the outlier detection methodology proposed by Li and Chan (2005, 2007) to find out the timings of the historical mortality outliers (jumps).[1]

[1] A detailed description about the data used in the outlier analysis is provided in Section 2.2. In implementing the outlier detection method, we consider positive additive outliers (i.e., outliers with no lasting impact) only, because the focus of this chapter is on short-term catastrophic mortality jumps.

Then for each detected mortality jump, we approximate its age pattern by computing

$$y(x,T^*) = \ln(m_{x,T^*}) - \frac{1}{6}\left(\sum_{t=T^*-3}^{T^*-1} \ln(m_{x,t}) + \sum_{t=T^*+1}^{T^*+3} \ln(m_{x,t})\right).$$

for all age group x, where T^* is the timing of the detected jump and $m_{x,t}$ is the central death rate for age group x at time t. This quantity compares the log death rate for each age group in the year when the mortality jump occurred with the corresponding average log death rate over the six neighboring years.[2] The patterns of $y(x,T^*)$ for all detected mortality jumps are depicted in Figure 2.1, from which we can conclude that the age patterns of mortality jumps are not uniform over age and exhibit certain degrees of variation. These properties cannot be reflected in models that are based on aggregate mortality indexes. Also shown in Figure 2.1 are the values of b_x (the age response parameters describing the age pattern of general mortality improvements) in the original Lee-Carter model (Lee and Carter, 1992) that is estimated to the data from each of the two populations.[3] It can be seen that the patterns of b_x and $y(x,T^*)$ are generally different, indicating that models using the same age response parameters for mortality jumps and general mortality improvements may not be adequate.

To our knowledge, the work of Cox et al. (2010) is the only attempt so far to explicitly address the age pattern of mortality jumps, but their modeling approach is based much more heavily on expert opinions than statistical estimation. To fill this gap, in this chapter we propose two variants of the Lee-Carter model with short-term jump effects. The first variant captures the age pattern of mortality jumps by a distinct collection of parameters, acknowledging the empirical fact that the age patterns of general and extreme changes in mortality rates over time are different. The second variant is a further generalization which permits the age pattern of mortality jumps to vary randomly, taking into account the correlation of jump effects among different age groups. Both model variants nest the transitory jump model developed by Chen and Cox (2009), in which mortality jumps are incorporated in the time-series process for the period effects. However, our proposed

[2]It can be shown easily that $y_{x,T^*} \approx J_{x,T^*}$ under our general model specification (equation (2.1)) and the random walk assumption (equation (2.2)), where J_{x,T^*} is a component in our model which measures the effect of a mortality jump occurred in year T^* on age group x. To prevent the potential masking effect arising from the noise in the data, we compare the mortality in year T^* with the average mortality over $T^* \pm 3$ years (rather than just $T^* \pm 1$ years). If another mortality jump occurred in the neighboring years, then the death rates for the year in which the other jump occurred are excluded in the calculation. For example, for the U.S. population, we compare the death rates in 1918 with the average death rates over years 1915-1917, 1919 and 1921. The death rates in year 1920, in which another jump occurred, are excluded in the calculation.

[3]The full definition of the original Lee-Carter model is provided in Section 2.3.

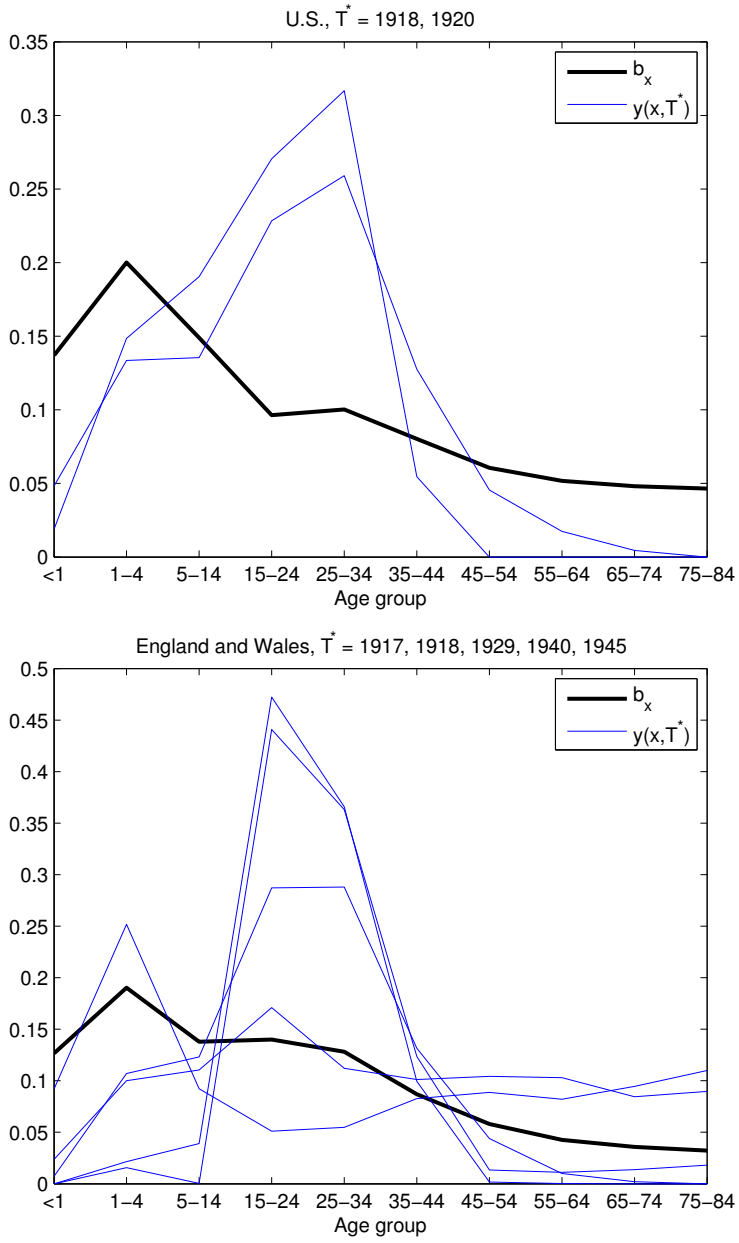


Figure 2.1: The estimates of $y(x, T^*)$ (the age patterns of mortality jumps) and b_x (the age response parameters in the original Lee-Carter model) for the populations of the U.S. (the upper panel) and England and Wales (the lower panel). To facilitate exposition, the values of $y(x, T^*)$ shown are normalized; i.e., the values of $y(x, T^*)/\sum_x y(x, T^*)$ are shown.

model variants demand different (and more advanced) estimation techniques, because in these model variants jump effects are involved in not only the time-series process but also other parts of the model structure.

We do acknowledge that parsimony is important and that the number of parameters grows as additional features are introduced. To focus on the issues we intend to investigate, we model one population only at a time and do not incorporate random changes in long-term mortality trends. Also, to preserve the tractability of the resulting log-likelihood function, we only consider Gaussian jumps, despite that other jump severity distributions such as double-exponential may produce a better fit. We overcome the challenges in statistical estimation by using the Route II estimation methodology that was recently introduced by Haberman and Renshaw (2012). This alternative estimation method is based on the first differences of the log mortality rates with respect to calendar time rather than the log mortality rates themselves. Compared to the traditional way of estimation, the advantages of the Route II methodology are twofold. First, in using the Route II approach, parameters representing the static level of mortality (i.e., the a_x parameters) are not involved in the estimation and projection processes.[4] By excluding these parameters, the model structure being estimated is more parsimonious and therefore convergence is easier to achieve. Second, in contrast to the traditional approach in which an extra step is needed to estimate the time-series process embedded in the model, the Route II methodology permits us to estimate all relevant parameters in one single estimation algorithm. So far as we aware, this thesis represents the first attempt to use the Route II approach to estimate stochastic mortality models with jump effects.

It is reasonable to conjecture that the features we consider have an impact of the pricing of catastrophic mortality bonds, because two mortality jumps of the same severity but different age patterns could affect the payout from a mortality bond differently, depending on the age range with which the mortality bond is associated. To verify this conjecture, we attempt to use our proposed model variants to price a collection of hypothetical mortality bonds that are associated with different age ranges. In an incomplete market, the pricing problem is not straightforward, although it can be accomplished by insurance-based methods (Chen and Cummins, 2010; Wills and Sherris, 2010), no-arbitrage methods (Cairns et al., 2006; Chen and Cox, 2009; Chen et al., 2013a; Li, 2010; Li and Ng, 2011; Lin and Cox, 2008) or economic methods (Zhou et al., 2011, 2013a,b; Chen et al., 2013b). The pricing method we use is the method of canonical valuation, which was first proposed by Stutzer (1996) and applied to the market of insurance-linked securities by Chen et al. (2013a), Li

[4]In using the Route II methodology, a projection of future mortality rates is obtained by first summing the projected first differences of future log mortality rates and then applying the summand to the most recent log mortality rates. The a_x parameters are therefore irrelevant.

(2010) and Li and Ng (2011). Using the Atlas IX Capital Class B note launched in 2013 as a martingale constraint, this method identifies a risk-neutral probability measure, from which prices of the hypothetical mortality bonds can be estimated.

The rest of this chapter is organized as follows. Section 2.2 describes the mortality data used in our illustrations. Section 2.3 provides the specifications of the proposed model variants. Section 2.4 explains how the parameters in the proposed model variants are estimated. Section 2.5 presents the estimation results and evaluates the goodness-of-fit. Section 2.6 details the pricing method we use and presents the calculated prices for a collection of hypothetical catastrophic mortality bonds. Finally, concluding remarks are provided in Section 2.7.

2.2 Mortality Data

The illustrations in this chapter are based on the mortality data from the unisex populations of the U.S. and England and Wales. We obtain the required data from two official sources.

The data for the U.S. population from 1901 to 2005 are provided by the Centers for Diseases Control and Prevention (CDC).[5] These data are arranged by age groups <1, 1-4, 5-14, 15-24, ..., 75-84.

For the U.S. population from 2006 to 2010 and English and Welsh population from 1901 to 2011, the data are obtained from the Human Mortality Database (2014). For consistency reasons, we group these data in the same way as in the data from the CDC.

2.3 Model Specification

2.3.1 The Original Lee-Carter Model

Our proposed model variants are built upon the original Lee-Carter model (Lee and Carter, 1992), which can be expressed as

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

where $m_{x,t}$ denotes the central death rate for age group x in year t. In the model, parameter a_x represents the static level of mortality for age group x, parameter k_t captures the

[5] www.cdc.gov

variation of log mortality rates over time, parameter b_x measures the sensitivity of $\ln(m_{x,t})$ to changes in k_t, and $e_{x,t}$ is the sampling error. It is assumed that $e_{x,t} \overset{iid}{\sim} N(0, \sigma_e^2)$.

It is well-known that the model is subject to an identifiability problem. To stipulate parameter uniqueness, the following identifiability constraints are often used:

$$\sum_{x=1}^{X} b_x = 1 \quad \text{and} \quad \sum_{t=t_1}^{t_T} k_t = 0,$$

where X is the number of age groups under consideration, $[t_1, t_T]$ is the data sample period and T is the length of the data sample period. These constraints are also used in this chapter.

To generate mortality forecasts, the evolution of k_t over time is modeled by a random walk with drift:

$$k_t = \mu + k_{t-1} + \xi_t, \tag{2.1}$$

where μ is the drift term and the innovation term ξ_t follows a normal distribution with mean 0 and variance σ^2. It is assumed that the innovations have no serial correlation and are independent of $e_{x,t}$. Given this process, mortality forecasts can be obtained readily by extrapolation.

2.3.2 The General Specification for the Model Variants under Consideration

We now present the general specification for all model variants we consider. To introduce short-term jump effects, we add an extra term $N_t J_{x,t}$ to the original Lee-Carter model, giving

$$\ln(m_{x,t}) = a_x + b_x k_t + N_t J_{x,t} + e_{x,t}, \tag{2.2}$$

where a_x, b_x and $e_{x,t}$ carry the same meanings as in the original Lee-Carter model. The sequence of k_t in equation (2.2) is assumed to be free of jumps, so a change in k_t can be interpreted to mean a general (but not extreme) change in the overall level of mortality. We retain the assumption that k_t follows the random walk with drift specified by equation (2.1).

We allow a maximum of one mortality jump in a calendar year. The jump count variable N_t equals one if a mortality jump occurs in year t, and zero otherwise. It follows

that N_t follows a Bernoulli distribution with parameter p, where p denotes the probability of a mortality jump in a calendar year. We assume that the sequence of N_t possesses no serial correlation.

The variable $J_{x,t}$ reflects the effect of a mortality jump occurred in year t on age group x. We assume Gaussian jumps, which imply the vector $\vec{J}_t = (J_{1,t}, \ldots, J_{X,t})'$ follows a multivariate normal distribution with mean vector $\vec{\mu}^{(J)}$ and covariance matrix Σ_J. We further assume that the sequence of $\vec{J}_t$ has no serial correlation and that $\vec{J}_t$ is independent of N_t, ξ_t and $e_{x,t}$.

In what follows we present three specific model variants. The differences among these model variants lie in the way in which $\vec{\mu}^{(J)}$ and Σ_J are specified.

2.3.3 Model J0

Model J0 is constructed by setting

$$\vec{\mu}^{(J)} = \vec{b}\mu_J \quad \text{and} \quad \Sigma_J = \begin{pmatrix} b_1^2 & b_1 b_2 & \cdots & b_1 b_X \\ b_1 b_2 & b_2^2 & & b_2 b_X \\ \vdots & & \ddots & \vdots \\ b_1 b_X & b_2 b_X & \cdots & b_X^2 \end{pmatrix} \sigma_J^2,$$

where $\vec{b} = (b_1, \ldots, b_X)'$. It follows that the model can be expressed as

$$\ln(m_{x,t}) = a_x + b_x(k_t + N_t Y_t) + e_{x,t},$$

where $Y_t \sim N(\mu_J, \sigma_J^2)$ is a random variable representing the severity of the jump occurring at time t. In Model J0, the jump effects $N_t Y_t$ are superimposed onto the general period effects k_t. The age patterns of jump effects and general period effects are assumed to be identical, determined by the same collection of b_x parameters.

Model J0 is not new and is in fact identical to the transitory jump model proposed by Chen and Cox (2009). It is used in this chapter to benchmark against our proposed model variants, in which the age pattern of mortality jumps is explicitly modeled.

We remark here that the way in which we estimate Model J0 is different from that used by Chen and Cox (2009). In this chapter, we use the Route II methodology proposed by Haberman and Renshaw (2012), in which all relevant parameters are estimated in a single estimation algorithm. By contrast, the estimation procedure of Chen and Cox (2009) involves multiple stages. First, they apply a rank-1 singular value decomposition to the

matrix of $\ln(m_{x,t}) - a_x$ to obtain estimates of b_x and k_t. Second, they re-estimate k_t so that for all t, the following equation holds:

$$D_t = \sum_{x=1}^{X}(E_{x,t}\exp(a_x + b_x k_t)),$$

where D_t is the total death count in year t and $E_{x,t}$ is the exposure count for age group x in year t. Then finally, the parameters associated with the jump effects and the random walk with drift are estimated by the method of conditional maximum likelihood. Their estimation method works for this model variant, but not for the next two model variants in which jump effects are involved in components other than the time-series process.

2.3.4 Model J1

Model J1 is constructed by setting

$$\vec{\mu}^{(J)} = \vec{b}^{(J)}\mu_J \quad \text{and} \quad \Sigma_J = \begin{pmatrix} (b_1^{(J)})^2 & b_1^{(J)}b_2^{(J)} & \cdots & b_1^{(J)}b_X^{(J)} \\ b_1^{(J)}b_2^{(J)} & (b_2^{(J)})^2 & & b_2^{(J)}b_X^{(J)} \\ \vdots & & \ddots & \vdots \\ b_1^{(J)}b_X^{(J)} & b_2^{(J)}b_X^{(J)} & \cdots & (b_X^{(J)})^2 \end{pmatrix} \sigma_J^2,$$

where $\vec{b}^{(J)} = (b_1^{(J)}, \ldots, b_X^{(J)})'$ is a constant vector. It immediately follows that Model J1 can be expressed as

$$\ln(m_{x,t}) = a_x + b_x k_t + b_x^{(J)} N_t Y_t + e_{x,t}. \tag{2.3}$$

In this model variant, the age pattern of jump effects is captured by a distinct parameter vector $\vec{b}^{(J)}$. It is clear that Model J1 nests Model J0, as the latter model can be recovered by setting $\vec{b}^{(J)}$ to $\vec{b}$. When estimating Model J1, an additional identifiability constraint

$$\sum_{x=1}^{X} b_x^{(J)} = 1$$

is used to stipulate parameter uniqueness. The use of this parameter constraint also permits us to readily compare the $b_x^{(J)}$ parameters with the b_x parameters, which also sum to one.

For this model variant, the age pattern of jump effects is fixed despite the jump severity Y_t is random. In more detail, equation (2.3) implies that the normalized effect of a mortality

jump occurred in year t on age group x is given by

$$\frac{J_{x,t}}{\sum_x J_{x,t}} = \frac{b_x^{(J)} Y_t}{\sum_x b_x^{(J)} Y_t} = b_x^{(J)},$$

which is non-random. The fact that Model J1 implies a fixed age pattern of mortality jumps can also be understood from the specification of Σ_J. In particular, the specification of Σ_J means that $J_{x,t}$ and $J_{y,t}$ for all $x, y = 1, \ldots, X$ are perfectly correlated with each other. We may therefore express $J_{x,t}$ as

$$J_{x,t} = b_x^{(J)}(\mu_J + \sigma_J W_t),$$

where W_t is a standard normal random variable that is independent of age. Therefore, no matter what the realized value of W_t is, the jump effects for age groups $1, \ldots, X$ must be proportional to $b_1^{(J)}, \ldots, b_X^{(J)}$.

2.3.5 Model J2

From Figure 2.1 (particularly in the lower panel) we observe that the (normalized) age patterns of different mortality jumps are somewhat different. It is therefore reasonable to further generalize Model J1 so that the normalized jump effect for age group x can deviate from $b_x^{(J)}$ in a random manner. To incorporate this property, one can permit the jump effects for different age groups to be correlated with one another, but not perfectly so.

In Model J2, we allow the age pattern of mortality jumps to vary by setting

$$\vec{\mu}^{(J)} = \vec{b}^{(J)} \mu_J \quad \text{and} \quad \Sigma_J = \begin{pmatrix} (b_1^{(J)})^2 & \rho b_1^{(J)} b_2^{(J)} & \cdots & \rho b_1^{(J)} b_X^{(J)} \\ \rho b_1^{(J)} b_2^{(J)} & (b_2^{(J)})^2 & & \rho b_2^{(J)} b_X^{(J)} \\ \vdots & & \ddots & \vdots \\ \rho b_1^{(J)} b_X^{(J)} & \rho b_2^{(J)} b_X^{(J)} & \cdots & (b_X^{(J)})^2 \end{pmatrix} \sigma_J^2,$$

where $-1 \leq \rho \leq 1$ is a constant, which can be interpreted to mean the correlation between the jump effects for age groups i and j, where $i \neq j$. It is clear that Model J2 nests Model J1, since the latter is the special case when $\rho = 1$. As before, the identifiability constraint $\sum_{x=1}^{X} b_x^{(J)} = 1$ is used when estimating Model J2.

This model variant permits the age pattern of jump effects to vary randomly, as the normalized jump effect $J_{x,t} / \sum_x J_{x,t}$ can no longer be simplified to a non-random constant.

From another viewpoint, the specification of Σ_J implies that

$$J_{x,t} = b_x^{(J)}(\mu_J + \sigma_J W_{x,t}),$$

where $W_{x,t} \sim N(0,1)$ and $\text{cov}(W_{x,t}, W_{y,t}) = \rho$ for $x \neq y$. It follows that when $x \neq y$, the realized values of $W_{x,t}$ and $W_{y,t}$ are different, which in turn means that the jump effects for age groups $1, \ldots, X$ are no longer proportional to $b_1^{(J)}, \ldots, b_X^{(J)}$.

We acknowledge that it is possible to further relax the model structure by permitting the correlations between different pairs of age groups to be different, that is, by having all $X(X-1)/2$ off-diagonal elements in Σ_J to be free parameters. However, adopting an unconstrained Σ_J matrix would significantly increase the number of parameters that need to be estimated. The simplified structure we use allows us to incorporate the potential variations in the age pattern of mortality jumps with only one additional parameter relative to Model J1.

2.4 Estimation Method

As previously mentioned, the estimation method of Chen and Cox (2009) is not applicable to Models J1 and J2, in which jump effects are not simply superimposed onto the random walk with drift. To overcome the estimation challenge, we use the Route II estimation methodology documented in the paper by Haberman and Renshaw (2012), which the authors found to work well under the Lee-Carter modeling framework.

The Route II estimation method is based on the first differences of the log mortality rates with respect to calendar time rather than the log mortality rates themselves. In more detail, the goal of this estimation method is to estimate the parameters contained in the structure of

$$Z_{x,t} := \ln(m_{x,t}) - \ln(m_{x,t-1}).$$

As $Z_{x,t} \approx \frac{\partial \ln(m_{x,t})}{\partial t} = \frac{1}{m_{x,t}} \frac{\partial m_{x,t}}{\partial t}$, we can interpret $Z_{x,t}$ to mean the scaled or relative mortality improvement rate at age x at time t.[6] An incremental mortality improvement implies $Z_{x,t} < 0$, while an incremental mortality deterioration implies $Z_{x,t} > 0$.

[6] There are alternative ways to approximate the partial derivative. For example, Haberman and Renshaw (2012) approximates it by using $Z_{x,t} = 2(1 - \frac{m_{x,t}}{m_{x,t-1}})/(1 + \frac{m_{x,t}}{m_{x,t-1}})$. We acknowledge that using $\Delta t = 1$ may not lead to an accurate approximation. This approximation is used solely to facilitate the demographic interpretation of $Z_{x,t}$ and is not used elsewhere in the chapter. All results in this chapter are independent of the accuracy of this approximation.

The general specification in equation (2.2) implies that

$$Z_{x,t} = b_x(k_t - k_{t-1}) + N_t \cdot J_{x,t} - N_{t-1} \cdot J_{x,t-1} + \varepsilon_{x,t}, \tag{2.4}$$

where $\varepsilon_{x,t} = e_{x,t} - e_{x,t-1}$. It is clear that $\varepsilon_{x,t} \sim N(0, \sigma_r^2)$, where $\sigma_r^2 = 2\sigma_e^2$. Under the assumption that k_t follows a random walk with drift, equation (2.2) can be further simplified into

$$Z_{x,t} = b_x(\mu + \xi_t) + N_t \cdot J_{x,t} - N_{t-1} \cdot J_{x,t-1} + \varepsilon_{x,t}. \tag{2.5}$$

For convenience, we rewrite equation (2.5) in vector form as follows:

$$\vec{Z}_t = \vec{b}(\mu + \xi_t) + N_t \vec{J}_t - N_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t, \tag{2.6}$$

where $\vec{Z}_t = (Z_{1,t}, \ldots, Z_{X,t})'$ and $\vec{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{X,t})'$.

Our goal is to estimate the parameters in the parametric structure of $\vec{Z}_t$. Noting that $\vec{Z}_t$ is not serially independent, we consider the method of conditional maximum likelihood, whereby the joint density of all observations is expressed as the product of the conditional densities of different observations. In particular, the log-likelihood function can be expressed as

$$\begin{aligned} l(\vec{\theta}) &= \ln f(\vec{z}_{t_2}, \ldots, \vec{z}_{t_T}; \vec{\theta}) \\ &= \ln[f(\vec{z}_{t_T}|\vec{z}_{t_{T-1}}; \vec{\theta}) \cdot f(\vec{z}_{t_{T-1}}|\vec{z}_{t_{T-2}}; \vec{\theta}) \ldots f(\vec{z}_{t_3}|\vec{z}_{t_2}; \vec{\theta}) \cdot f(\vec{z}_{t_2}; \vec{\theta})] \\ &= \ln f(\vec{z}_{t_2}; \vec{\theta}) + \sum_{t=t_3}^{t_T} \ln f(\vec{z}_t|\vec{z}_{t-1}; \vec{\theta}) \\ &= \sum_{t=t_2}^{t_{T-1}} \ln f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta}) - \sum_{t=t_3}^{t_{T-1}} \ln f(\vec{z}_t; \vec{\theta}), \end{aligned}$$

where $\vec{\theta}$ denotes the vector of parameters being estimated. The second step in the derivation above originates from the fact that $\vec{Z}_t$ depends on $\vec{Z}_{t-1}$ but not $\vec{Z}_s$ for any integer $s < t-1$.

On the basis of the distributional assumptions we imposed on ξ_t, N_t, $\vec{J}_t$ and $\vec{\varepsilon}_t$, it can be shown that the following three properties hold[7]:

Property 1

$$\vec{Z}_t|N_t = n_t, \vec{J}_t = \vec{j}_t \quad \perp \quad \vec{Z}_{t+1}|N_t = n_t, \vec{J}_t = \vec{j}_t.$$

[7]The derivation of the three properties is summarized in Appendix A.1.

Property 2

$$\vec{Z}_t | N_{t-1} = n_{t-1}, N_t = n_t \sim MVN(\vec{\mu}_{mar}, \Sigma_{mar}),$$

where

$$\vec{\mu}_{mar} = \vec{b}\mu + (n_t - n_{t-1})\vec{\mu}^{(J)}$$

and

$$\Sigma_{mar} = \vec{b}\vec{b}'\sigma^2 + \mathrm{I} \cdot \sigma_r^2 + (n_t^2 + n_{t-1}^2)\Sigma_J.$$

Property 3 Define

$$\vec{Z}^*_{t,t+1} = \begin{pmatrix} \vec{Z}_t \\ \vec{Z}_{t+1} \end{pmatrix}.$$

We have

$$\vec{Z}^*_{t,t+1} | N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1} \sim MVN(\vec{\mu}_{joint}, \Sigma_{joint}),$$

where

$$\vec{\mu}_{joint} = \begin{pmatrix} \vec{b}\mu + (n_t - n_{t-1})\vec{\mu}^{(J)} \\ \vec{b}\mu + (n_{t+1} - n_t)\vec{\mu}^{(J)} \end{pmatrix}$$

and

$$\Sigma_{joint} = \left(\begin{array}{c|c} \sigma^2\vec{b}\vec{b}' + \mathrm{I} \cdot \sigma_r^2 + (n_t^2 + n_{t-1}^2)\Sigma_J & -n_t^2\Sigma_J \\ \hline -n_t^2\Sigma_J & \sigma^2\vec{b}\vec{b}' + \mathrm{I} \cdot \sigma_r^2 + (n_{t+1}^2 + n_t^2)\Sigma_J \end{array} \right).$$

Using these three properties, we can calculate the unconditional density $f(\vec{z}_t; \vec{\theta})$ and the joint density $f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})$ in the log-likelihood function by first conditioning on the jump indicator variables and then integrating the conditional density over the jump indicator variables. Specifically, the unconditional density of $\vec{z}_t$ can be expressed as

$$f(\vec{z}_t; \vec{\theta}) = \sum_{n_t=0}^{1} \sum_{n_{t+1}=0}^{1} f(\vec{z}_t | N_{t-1} = n_{t-1}, N_t = n_t) \Pr(N_{t-1} = n_{t-1}, N_t = n_t),$$

whereas the joint density of $\vec{z}_t$ and $\vec{z}_{t+1}$ is given by

$$\begin{aligned} f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta}) &= \sum_{n_t=0}^{1} \sum_{n_{t+1}=0}^{1} \sum_{n_{t+2}=0}^{1} f(\vec{z}_t, \vec{z}_{t+1} | N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1}) \\ &\quad \times \Pr(N_t = n_t, N_{t+1} = n_{t+1}, N_{t+2} = n_{t+2}). \end{aligned}$$

The log-likelihood function $l(\vec{\theta})$ is maximized by an iterative Newton-Raphson procedure, in which parameters are updated one at a time. The details about the iterative procedure, including the initial values used, are provided in Appendix A.2.

Given the parameter estimates, one can readily obtain a projection of future mortality rates by first summing the projected first differences of future log mortality rates and then applying the summand to the most recent log mortality rates. In particular, the projected value of $\ln(m_{x,u})$ for $u = t_{T+1}, t_{T+2}, \ldots$ can be computed by adding the projected value of $\sum_{t=t_{T+1}}^{u} Z_{x,t}$ to $\ln(m_{x,t_T})$.

We conclude this section with three remarks. First, since $Z_{x,t}$ does not depend on a_x, the a_x parameters are not estimated. The irrelevance of the a_x parameters is not surprising, because in using the Route II approach, the static level of mortality is determined by the log mortality rates at the forecast origin, instead of the a_x parameters. By excluding these irrelevant parameters, the parametric structure being estimated is more parsimonious and hence convergence can be achieved more easily. Second, the Route II projection approach, i.e., applying projected $Z_{x,t}$ values to the value of $\ln(m_{x,t_T})$, ensures that the projected mortality rates progress naturally from the mortality rates at the forecast origin, thereby avoiding the potential short-term forecast bias that may result from the typical two-stage estimation method with $\sum_{x=1}^{X} b_x = 1$ and $\sum_{t=t_1}^{t_T} k_t = 0$ as identifiability constraints (see Lee and Miller, 2001; Li et al., 2009). This advantage is particularly important when the modeler's objective is to value catastrophic mortality bonds, which often have short maturities. Third, the Route II approach allows us to estimate all relevant parameters, including the parameters in the random walk with drift, in one single estimation algorithm. This alternative estimation approach therefore spares us from the problems, for example, the need for standard error corrections, that are associated with multi-stage estimation methods (see Murphy and Topel, 2002).

2.5 Estimation Results

All three model variants are fitted to the historical mortality data from the U.S. and English and Welsh populations. The estimated parameters and their standard errors are shown in Tables 2.1 and 2.2. The standard errors are calculated by a parametric bootstrapping

procedure, which is detailed in Appendix A.3. Admittedly, some of the parameter estimates that are associated with jump effects have rather large standard errors. The large standard errors are not overly surprising, because over the data sample period there were only a handful of extreme changes in mortality on which the estimation of these parameters could be based.

Because the three model variants are nested, we can readily compare their fit to the historical data by using the Akaike Information Criterion (AIC) (Akaike, 1974), which is defined by $2 \times (-\hat{l} + k)$, where $\hat{l}$ is the maximized log-likelihood and k is the effective number of parameters. We prefer the model with the smallest AIC value, which indicates the model provides the best fit (measured by $\hat{l}$), taken into account its degree of parsimony (measured by k). The AIC values of all models we estimated are displayed in Tables 2.1 and 2.2.

For both populations, Model J1 gives a lower AIC value in comparison to Model J0. This result indicates that the fit to historical data is improved significantly if a distinct collection of parameters is used to capture the age pattern of mortality jumps. For the population of England and Wales, Model J2 gives an even smaller AIC value, which means that the fit to historical data is further benefited by allowing the age pattern of mortality jumps to vary randomly. However, this conclusion does not apply to the U.S. population, possibly because the U.S. population were subject to fewer extreme changes in mortality over the data sample period and thus the information contained in the historical data might not be sufficient to support the use of a varying age pattern. On the basis of the AIC value, we recommend Model J1 for the U.S. population and Model J2 for English and Welsh population.

We can also conduct the likelihood ratio test to compare the three model variants. The null hypothesis and alternative hypothesis of the test are summarized as follow:

$$H_0: \quad \vec{\theta} = \vec{\theta}_0 \quad \text{and} \quad H_1: \quad \vec{\theta} = \vec{\theta}_1,$$

where $\vec{\theta}_0$ represents the set of parameters from the unrestricted model, and $\vec{\theta}_1$ represents the set of parameters from the restricted model. The likelihood ratio test statistic LR is calculated as

$$LR = 2 \times (\hat{l}_1 - \hat{l}_0),$$

where $\hat{l}_0$ is the maximized log-likelihood of the unrestricted model and $\hat{l}_1$ is the maximized log-likelihood of the restricted model. The results of the likelihood ratio test for both populations are summarized in Table 2.3. Same as comparing the AIC value, the likelihood ratio test results suggest Model J1 for the U.S. population and Model J2 for English and

Table 2.1: Estimates of the parameters in Models J0, J1 and J2, the U.S. unisex population. The standard errors of the parameter estimates are shown in parentheses.

Estimates (standard error)								
Model J0								
b_1	0.071817	(0.005502)	μ	−0.159414	(0.024615)			
b_2	0.144984	(0.005706)	σ	0.230318	(0.021654)			
b_3	0.125876	(0.005514)	σ_r	0.035369	(0.000798)			
b_4	0.174212	(0.005599)	μ_J	0.322572	(1.182600)			
b_5	0.189691	(0.005681)	σ_J	1.988527	(0.696490)			
b_6	0.115635	(0.005483)	p	0.047808	(0.026661)			
b_7	0.063679	(0.005555)						
b_8	0.045722	(0.005594)						
b_9	0.038553	(0.005550)						
b_{10}	0.029831	(0.005592)						
Log-likelihood:	1963.148772		AIC:	−3896.297543				
Model J1								
b_1	0.104906	(0.004947)	$b_1^{(J)}$	0.045360	(0.008899)	μ	−0.203843	(0.044046)
b_2	0.161054	(0.005133)	$b_2^{(J)}$	0.133287	(0.006246)	σ	0.452523	(0.031716)
b_3	0.114020	(0.005017)	$b_3^{(J)}$	0.135024	(0.006391)	σ_r	0.026349	(0.000606)
b_4	0.113396	(0.005047)	$b_4^{(J)}$	0.222280	(0.013361)	μ_J	0.740631	(0.652480)
b_5	0.109291	(0.005043)	$b_5^{(J)}$	0.253698	(0.016476)	σ_J	1.362288	(0.421620)
b_6	0.097629	(0.004887)	$b_6^{(J)}$	0.129772	(0.006698)	p	0.064164	(0.026342)
b_7	0.080788	(0.005020)	$b_7^{(J)}$	0.049746	(0.007012)			
b_8	0.074319	(0.004995)	$b_8^{(J)}$	0.022695	(0.008335)			
b_9	0.071807	(0.004949)	$b_9^{(J)}$	0.012077	(0.008511)			
b_{10}	0.072788	(0.005018)	$b_{10}^{(J)}$	−0.003939	(0.010254)			
Log-likelihood	2188.382845		AIC:	−4324.765690				
Model J2								
b_1	0.104338	(0.004922)	$b_1^{(J)}$	0.046579	(0.006603)	μ	−0.203220	(0.043805)
b_2	0.160182	(0.004987)	$b_2^{(J)}$	0.133484	(0.005584)	σ	0.456687	(0.032294)
b_3	0.114087	(0.004945)	$b_3^{(J)}$	0.136716	(0.004888)	σ_r	0.026059	(0.000603)
b_4	0.114347	(0.004906)	$b_4^{(J)}$	0.220386	(0.009940)	μ_J	0.711151	(0.814950)
b_5	0.109416	(0.004985)	$b_5^{(J)}$	0.250450	(0.013630)	σ_J	1.891909	(0.537790)
b_6	0.098078	(0.004897)	$b_6^{(J)}$	0.128856	(0.004895)	p	0.071365	(0.027124)
b_7	0.080863	(0.004907)	$b_7^{(J)}$	0.050046	(0.005224)	ρ	0.999070	(0.000845)
b_8	0.074234	(0.004918)	$b_8^{(J)}$	0.023430	(0.006141)			
b_9	0.071752	(0.004866)	$b_9^{(J)}$	0.012904	(0.006177)			
b_{10}	0.072704	(0.004937)	$b_{10}^{(J)}$	−0.002852	(0.008445)			
Log-likelihood:	2188.921027		AIC:	−4323.842054				

Table 2.2: Estimates of the parameters in Models J0, J1 and J2, English and Welsh unisex population. The standard errors of the parameter estimates are shown in parentheses.

Estimates (standard error)								
Model J0								
b_1	0.018286	(0.005950)	μ	−0.113471	(0.017376)			
b_2	0.105161	(0.005731)	σ	0.092098	(0.031653)			
b_3	0.095571	(0.005837)	σ_r	0.067778	(0.001547)			
b_4	0.330116	(0.007189)	μ_J	1.632288	(0.657920)			
b_5	0.266008	(0.006752)	σ_J	1.967868	(0.443460)			
b_6	0.107817	(0.005889)	p	0.118019	(0.035530)			
b_7	0.038190	(0.005870)						
b_8	0.019531	(0.005868)						
b_9	0.014950	(0.005941)						
b_{10}	0.004369	(0.006004)						
Log-likelihood:	1313.331494		AIC:	−2596.662988				
Model J1								
b_1	0.112529	(0.008572)	$b_1^{(J)}$	−0.001796	(0.005827)	μ	−0.203858	(0.046202)
b_2	0.256623	(0.009615)	$b_2^{(J)}$	0.072235	(0.008187)	σ	0.459819	(0.034251)
b_3	0.152161	(0.008274)	$b_3^{(J)}$	0.080577	(0.004546)	σ_r	0.043686	(0.001013)
b_4	0.066560	(0.008935)	$b_4^{(J)}$	0.389606	(0.013685)	μ_J	1.998496	(0.410110)
b_5	0.072099	(0.008844)	$b_5^{(J)}$	0.308119	(0.010146)	σ_J	1.320538	(0.274390)
b_6	0.086473	(0.008218)	$b_6^{(J)}$	0.110848	(0.003734)	p	0.133860	(0.035408)
b_7	0.078347	(0.008302)	$b_7^{(J)}$	0.028514	(0.004050)			
b_8	0.060935	(0.008351)	$b_8^{(J)}$	0.010598	(0.004232)			
b_9	0.057952	(0.008394)	$b_9^{(J)}$	0.006409	(0.004294)			
b_{10}	0.056321	(0.008559)	$b_{10}^{(J)}$	−0.005110	(0.004481)			
Log-likelihood	1646.198340		AIC:	−3240.396680				
Model J2								
b_1	0.115010	(0.008248)	$b_1^{(J)}$	0.003626	(0.002792)	μ	−0.202446	(0.045380)
b_2	0.257910	(0.009592)	$b_2^{(J)}$	0.081977	(0.003841)	σ	0.453815	(0.033916)
b_3	0.149043	(0.008348)	$b_3^{(J)}$	0.086463	(0.002171)	σ_r	0.042855	(0.001006)
b_4	0.068967	(0.008840)	$b_4^{(J)}$	0.375667	(0.006481)	μ_J	1.133851	(1.439400)
b_5	0.066487	(0.008902)	$b_5^{(J)}$	0.293454	(0.005001)	σ_J	3.927853	(0.865820)
b_6	0.083765	(0.008332)	$b_6^{(J)}$	0.109057	(0.001901)	p	0.131646	(0.034135)
b_7	0.078472	(0.008308)	$b_7^{(J)}$	0.031342	(0.002014)	ρ	0.999510	(0.000012)
b_8	0.062099	(0.008336)	$b_8^{(J)}$	0.012886	(0.002021)			
b_9	0.059538	(0.008263)	$b_9^{(J)}$	0.008506	(0.002044)			
b_{10}	0.058710	(0.008477)	$b_{10}^{(J)}$	−0.002979	(0.002136)			
Log-likelihood:	1648.815577		AIC:	−3243.631154				

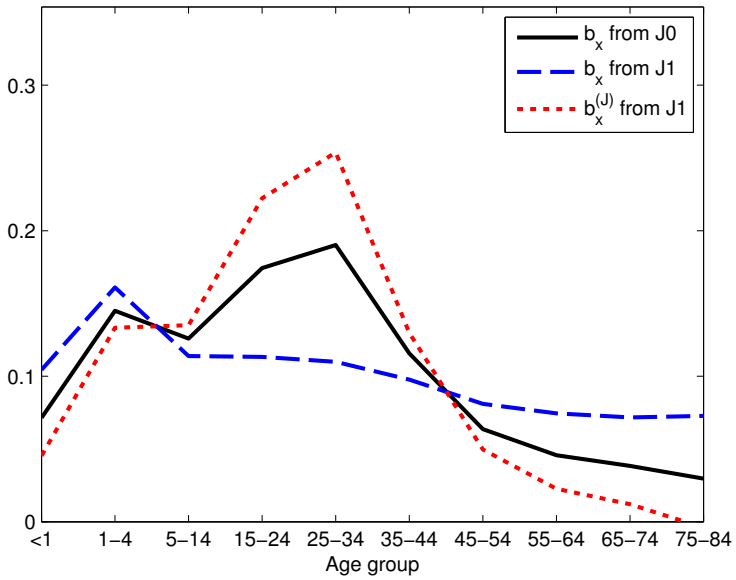
Welsh population.

Table 2.3: The likelihood ratio test results for the U.S. unisex population and English and Welsh unisex population.

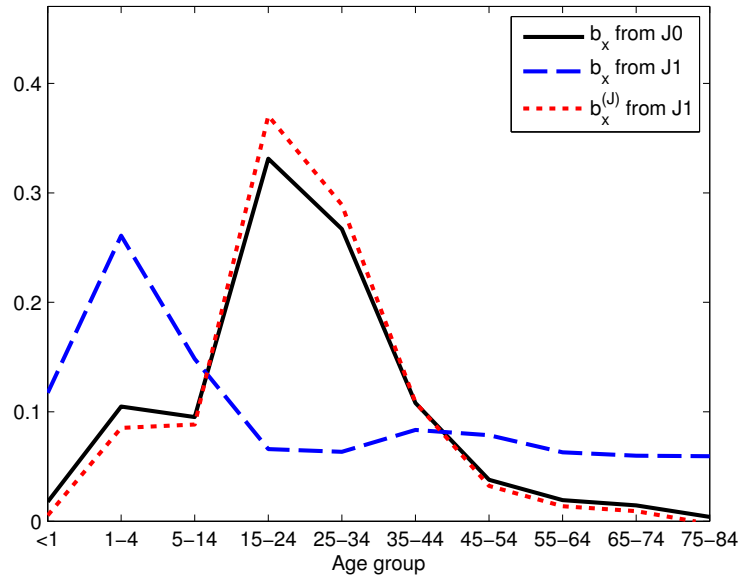
the U.S. unisex population				
$\vec{\theta}_0$	$\vec{\theta}_1$	LR	Critical Value	p-value
Model J0	Model J1	450.468146	18.307038	0.000000
Model J1	Model J2	1.076364	3.841459	0.299512
English and Welsh unisex population				
$\vec{\theta}_0$	$\vec{\theta}_1$	LR	Critical Value	p-value
Model J0	Model J1	665.733692	18.307038	0.000000
Model J1	Model J2	5.234474	3.841459	0.022144

In Figure 2.2 we compare the estimates of b_x in Model J0 and the estimates of b_x and $b_x^{(J)}$ in Model J1/J2. Because the same normalization scheme applies to all three sets of parameters, we can gauge their relative importance readily. In Model J1/J2, the age pattern of a typical mortality jump is captured by the pattern of $b_x^{(J)}$, which has a peak in age group 15-24 / 25-34 and diminishes quickly towards the ends of the age spectrum, while the age pattern of general mortality improvements is captured by the pattern of b_x, which has a peak in age group 1-4 and is more uniform across different age groups. By contrast, in Model J0, the age patterns of extreme and general changes in mortality are both incorporated into the pattern of b_x, which looks like an average between the patterns of b_x and $b_x^{(J)}$ in Model J1/J2. The potential problems of Model J0 have now become clear. For example, the effect of a mortality jump on individuals aged between 15 and 34 could be understated, while that on individuals aged beyond 45 could be overstated.

In Figure 2.3 we display the normalized age patterns of 50 mortality jumps that are simulated from the estimated Models J1 and J2 for English and Welsh population. As expected, the normalized age patterns of all 50 simulated mortality jumps from Model J1 are identical. The fixed normalized age pattern is not adequate to capture what we found from the exploratory analysis for this population (see Section 2.1): the age patterns of the detected historical mortality jumps for English and Welsh population are quite different from one anther; out of the five detected jumps, three have a peak in age group 15-24, one has a peak in age group 1-4 and one has a peak in age group 25-34. By contrast, the jumps simulated from Model J2 appear to have a close resemblance to the historical jumps that are shown in the lower panel of Figure 2.1.



The U.S. population



English and Welsh population

Figure 2.2: The estimated values of b_x in Model J0 and $b_x^{(J)}$ in Model J1 (for the U.S. population) and Model J2 (for English and Welsh population).

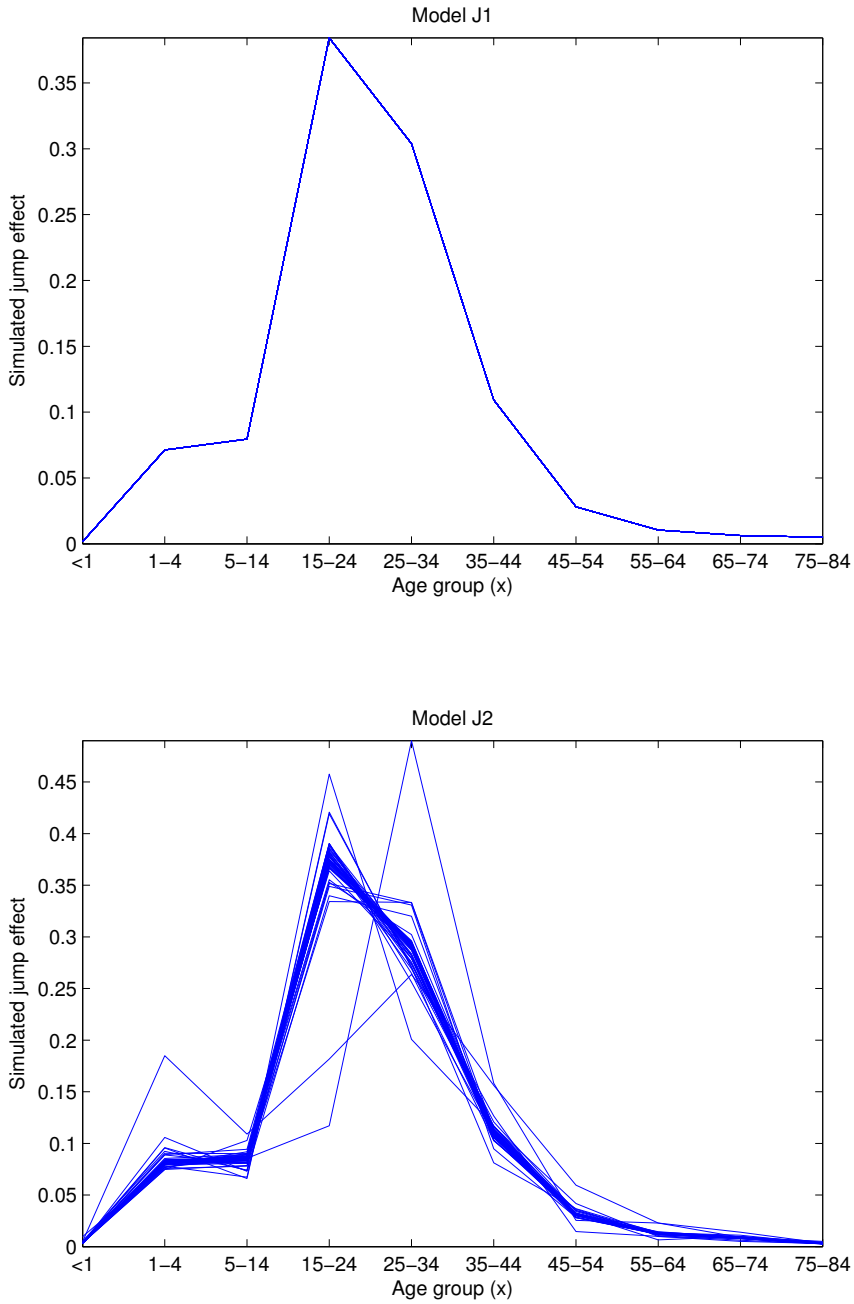


Figure 2.3: 50 mortality jumps simulated from Models J1 and J2 for the population of England and Wales. To facilitate exposition, the values shown are normalized; i.e., the magnitudes of $J_{x,t}/\sum_x J_{x,t}$ are shown.

2.6 Applications to Catastrophic Mortality Bonds Pricing

2.6.1 Catastrophic Mortality Bonds

In recent years, several reinsurers have attempted to cede their exposures to extreme mortality risk by issuing catastrophic mortality bonds. These bonds have rather short times to maturity, usually three to five years. Coupon payments are typically linked to a market interest rate such as the London Interbank Offered Rate (LIBOR), while the principal repayment is not guaranteed. In particular, the principal would be reduced if the underlying index exceeds an attachment point and exhausted if the index exceeds an exhaustion point. A typical underlying index is based on a weighted average of mortality rates across different populations, age groups and genders. The exact composition of it depends on the hedger's (seller's) risk exposures.

To attract investors with different objectives and risk preferences, a catastrophic mortality bond may be divided into several tranches, each of which has its own attachment and exhaustion points, coupon rates, credit rating and/or time to maturity. For example, six different tranches have been issued under Swiss Re's Vita Capital IV program that was launched in October 2010.[8] We refer interested readers to Bauer and Kramer (2009) and Blake et al. (2013) for more information about previously issued catastrophic mortality bonds.

2.6.2 Pricing Methodology

The market for catastrophic mortality bonds is incomplete in the sense that it is not possible to price a security in this market by constructing a replicating portfolio. In an incomplete market, a critical step in performing risk-neutral valuation is to identify a risk-neutral probability measure, under which prices of mortality bonds can be computed. Although this step can be accomplished by methods such as the Wang transform (see, e.g., Chen and Cox, 2009) and the Esscher transform (see, e.g., Chuang and Brockett, 2014), we consider the method of canonical valuation, which can be implemented readily with the model variants we consider.

The method of canonical valuation was introduced by Stutzer (1996). In his original work, valuation was performed in a fully non-parametric manner. Specifically, he derived

[8] See http://www.artemis.bm/deal_directory/vita-capital-iv-ltd/ and http://www.artemis.bm/deal_directory/vita-capital-iv-ltd-series-v-and-vi/.

a risk-neutral probability measure by adjusting the empirical distribution of future index values that was obtained by sampling with replacement from historical index values. The approach we consider is a semi-parametric variant (Li, 2010; Chen et al., 2013a), whereby risk-neutralization is applied to the empirical distribution of the sample paths generated from an assumed model.

The first step in implementing this pricing method is to generate a large number of sample paths of future mortality rates from the assumed stochastic mortality model, defined under the real-world probability measure. The generated sample paths represent as a collection of states of nature, all of which are equally probable. Therefore, if N sample paths are generated, then the probability mass function for the state of nature ω under the real-world probability measure $\mathbb{P}$ is given by

$$\Pr(\omega = \omega_j) = \pi_j = \frac{1}{N}, \qquad j = 1, 2, \ldots, N.$$

Our goal is to determine the probability distribution of ω under a risk-neutral probability measure $\mathbb{Q}$ that is equivalent to $\mathbb{P}$. We use $N = 20000$ in our calculations.

Let us suppose that the market contains m distinct primary securities, whose values evolve according to the state of nature ω. The ith primary security, where $i = 1, 2, \ldots, m$, has a time-0 price of F_i and, at the risk-free interest rate, a random discounted payoff of $f_i(\omega)$. We require that all primary securities are priced correctly under $\mathbb{Q}$, so that the martingale constraints

$$E^{\mathbb{Q}}[f_i(\omega)] = \sum_{j=1}^{N} f_i(\omega)\pi_j^* = F_i, \qquad i = 1, 2, \ldots, m, \tag{2.7}$$

where π_j^*, $j = 1, 2, \ldots, N$, is the probability distribution of ω under $\mathbb{Q}$, are satisfied.

In an incomplete market, $m < N$ and hence there are multiple risk-neutral probability measures satisfying (2.7). Let $\mathcal{Q}$ be the set of all measures that are equivalent to $\mathbb{P}$ and satisfy (2.7); that is, $\mathcal{Q}$ is the set of all equivalent martingale measures. The next step is to choose a measure in $\mathcal{Q}$. This choice is based on the Kullback-Leibler information criterion (Kullback and Leibler, 1951), which is defined by

$$D(\mathbb{Q}, \mathbb{P}) = E^{\mathbb{P}}\left(\frac{d\mathbb{Q}}{d\mathbb{P}} \ln \frac{d\mathbb{Q}}{d\mathbb{P}}\right) = \sum_{j=1}^{N} \pi_j^* \ln \frac{\pi_j^*}{\pi_j}.$$

We choose the equivalent martingale measure $\mathbb{Q}_0$ that minimizes the Kullback-Leibler

information criterion; that is,

$$\mathbb{Q}_0 = \arg\min_{\mathbb{Q}\in\mathcal{Q}} D(\mathbb{Q},\mathbb{P}),$$

subject to $\sum_{j=1}^{N}\pi_j^* = 1$ and the martingale constraints specified by equation (2.7). We refer $\mathbb{Q}_0$ to as the canonical measure.

From a statistical angle, the justification for the canonical measure is that it incorporates all information contained in the prices of the m primary securities that are traded in the market but no other (irrelevant) information. The canonical measure can also be justified from economic and geometric viewpoints (see Frittelli, 2000; Li and Ng, 2011).

Given the canonical measure, we can readily price a security that is associated with the same underlying. Let us consider a security that has a payoff, discounted to time-zero at the risk-free interest rate, of $g(\omega_j)$ in the jth state of nature. The price of this security implied by $\mathbb{Q}_0$ is simply $\sum_{j=1}^{N} g(\omega_j)\tilde{\pi}_j^*$, where $\tilde{\pi}_j^*$, $j = 1, 2, \ldots, N$, is the probability distribution of ω under Q_0.

2.6.3 Derivation of the Canonical Measure

We consider $m = 1$ martingale constraint, which is based on the price (premium spread) of the Atlas IX Capital Class B note launched by SCOR Global Life in 2013. The note is designed to protect the issuer from losses that are associated with extreme increases in the U.S. mortality over a six-year observation period from January 1, 2013 to December 31, 2018. It makes coupon payments that are linked to the 3-month LIBOR quarterly and a principal repayment at maturity. The principal would be reduced if there is an increase in the underlying U.S. mortality index that exceeds a specified percentage of a reference mortality index during the observation period. The key information about this note is summarized in Table 2.4.[9]

The actual underlying index is composed by a weighted average of U.S. mortality rates over different age groups and genders. In our valuation work, we use a slightly simplified index that is based on unisex mortality rates, because the models we consider are fitted to unisex rather than gender-specific mortality. Each age weight in the simplified index is the sum of the corresponding age weights for both genders in the actual index.[10] The age weights used in the simplified mortality index are displayed in Table 2.5. On the basis

[9]Source: S&P Rating Report (www.standardandpoors.com/ratings/ils/en/us)

[10]The age and gender weights used in the actual underlying mortality index can be found in the S&P Rating Report (www.standardandpoors.com/ratings/ils/en/us).

Table 2.4: Information about the variable-rate principal-at-risk series 2013-1 Class B notes issued under the Atlas IX Capital program.

Credit rating	BB
Amount	$180 million
Interest	3-month LIBOR minus 6 bps (floored at zero) plus 325 bps
Term	5 years and 4 months
Closing date	September 11, 2013
Scheduled maturity date	January 17, 2019

Table 2.5: The age weights used in the simplified mortality index.

Age Group (x)		Weight (w_x)
1	(Ages <1)	0.0000%
2	(Ages 1 - 4)	0.0000%
3	(Ages 5 - 14)	0.0000%
4	(Ages 15 - 24)	0.6479%
5	(Ages 25 - 34)	6.8844%
6	(Ages 35 - 44)	29.0933%
7	(Ages 45 - 54)	34.0283%
8	(Ages 55 - 64)	19.6846%
9	(Ages 65 - 74)	6.8641%
10	(Ages 75 - 84)	2.7974%
Total		100.0000%

of these age weights, the simplified mortality index q_t for years $t = 2012, \ldots, 2018$ can be calculated as

$$q_t = \sum_{x=1}^{10} w_x m_{x,t}, \tag{2.8}$$

where w_x represents the weight on age group x.

The principal of this note is subject to a potential reduction at the end of each year from 2014 to 2018. In particular, at the ends of years $t = 2014, \ldots, 2018$, the mortality improvement index, $q_t^{(\mathrm{MI})}$, is calculated according to the following equation:

$$q_t^{(\mathrm{MI})} = \frac{q_{t-1} + q_t}{2 \times q_{t-2}^{(\mathrm{ref})}}, \tag{2.9}$$

where $q_t^{(\text{ref})}$ is the reference index, which is defined as follows:

$$q_t^{(\text{ref})} = \begin{cases} L \times q_{t-1}^{(\text{ref})}, & \text{if} \quad L \times q_{t-1}^{(\text{ref})} \leq q_t^{(\text{ref})} \\ (2-L) \times q_{t-1}^{(\text{ref})}, & \text{if} \quad q_t^{(\text{ref})} \leq (2-L) \times q_{t-1}^{(\text{ref})} \\ q_t^{(\text{ref})}, & \text{if} \quad (2-L) \times q_{t-1}^{(\text{ref})} < q_t^{(\text{ref})} < L \times q_{t-1}^{(\text{ref})} \end{cases}, \quad (2.10)$$

for $t = 2013, \cdots, 2016$, where $q_{2012}^{(\text{ref})} = q_{2012}$ and L denotes the attachment point. We can interpret $q_t^{(\text{MI})}$ to mean the change in mortality (on a two-year aggregate basis) relative to the reference index. The percentage reduction in the original principal at the ends of years $t = 2014, \ldots, 2018$ is a function of $q_t^{(\text{MI})}$:

$$R_t = \frac{\min(q_t^{(\text{MI})}, U) - \min(q_t^{(\text{MI})}, L)}{U - L}, \quad (2.11)$$

where $L = 1.02$ and $U = 1.04$ are the attachment and exhaustion points, respectively. Let P_t be the balance of the principal (per \$1 of the original principal) in year t. We have $P_{2014} = 1$ and

$$P_t = \max\left(100\% - \sum_{u=2014}^{t-1} R_u, 0\right) \quad (2.12)$$

for $t = 2015, \ldots, 2019$.

At the end of each quarter, a coupon is paid on the balance of the principal. Let C_t be the coupon (per \$1 of the original principal) payable at time t and c be the coupon rate. We have

$$C_{t+\frac{s}{4}} = cP_t,$$

for $t = 2014, \ldots, 2018$ and $s = 0, 1, 2, 3$, and $C_{2019} = cP_{2019}$. The coupon rate c is set to be the 3-month LIBOR minus 6 basis points (floored at zero) plus 325 basis points. As the proceeds from the sale of the notes are deposited into a designated collateral account that earns the 3-month LIBOR minus 6 basis points, the additional 325 basis points can be regarded as the premium spread that compensates investors for taking on the extreme mortality risk.

On the basis of the Atlas IX Capital Class B note, the martingale constraint for deriving

the canonical measure can be expressed as

$$E^{\mathbb{Q}}\left(\sum_{t=0}^{4}\sum_{s=0}^{3} D_{t+\frac{s+1}{4}} C_{2014+t+\frac{s}{4}} + D_{5+\frac{1}{4}}(C_{2019}+P_{2019})\right) = 1, \tag{2.13}$$

where D_t represents the risk-free discount factor for a period of t years. Let $V(\omega_j)$ be the value of $\sum_{t=0}^{4}\sum_{s=0}^{3} D_{t+\frac{s+1}{4}} C_{2014+t+\frac{s}{4}} + D_{5+\frac{1}{4}}(C_{2019}+P_{2019})$ in the jth state of nature (simulated mortality scenario). It can be shown that the distribution of ω under the resulting canonical measure can be expressed as

$$\tilde{\pi}_j^* = \frac{\exp(\hat{\gamma}V(\omega_j))}{\sum_{j=1}^{N}\exp(\hat{\gamma}V(\omega_j))}, \quad j = 1, 2, \ldots, N,$$

where the Lagrangian multiplier $\hat{\gamma}$ is given by

$$\hat{\gamma} = \arg\min_{\gamma} \sum_{j=1}^{N} \exp(\gamma(V(\omega_j) - 1)).$$

In our calculations, the simulated mortality scenarios are obtained by applying the simulated values of $Z_{x,t}$ (from either one of the three model variants) to the 2010 mortality rates,[11] assuming that they are identical to the 2012 mortality rates on which the first reference index value is based. We believe that the effect of this two-year time lag is not that material, as the principal reduction is based on relative changes in mortality rather than absolute levels of mortality. The risk-free discount factors are computed by using the yield curve for zero-coupon U.S. Treasury bonds and notes on September 11, 2013 (the closing date).[12]

2.6.4 Pricing Hypothetical Mortality Bonds

We consider three hypothetical mortality bonds that are linked to mortality indexes with different age weights. The age weights (i.e., w_x for $x = 1, \ldots, 10$) on which the mortality indexes are based are shown in Table 2.6. The first set of age weights is the same as that used in the (simplified) mortality index to which the Atlas IX Capital Class B note is linked. The other two sets of age weights, which are more concentrated on younger and

[11] The end of the data sample period is 2010.

[12] The yields on zero-coupon U.S. Treasury bonds and notes for selected maturities are provided by *Bloomberg*. The yields for other maturities are obtained by linear interpolations.

Table 2.6: The age weights used in the mortality indexes to which the hypothetical mortality bonds being valued are linked.

Age group (x)		Weight (w_x)		
		Hypothetical bond 1	Hypothetical bond 2	Hypothetical bond 3
1	(Age < 1)	0.00 %	0.00 %	0.00 %
2	(Ages 1 - 4)	0.00 %	0.00 %	0.00 %
3	(Ages 5 - 14)	0.00 %	10.00 %	0.00 %
4	(Ages 15 - 24)	0.65 %	40.00 %	0.00 %
5	(Ages 25 - 34)	6.88 %	40.00 %	0.00 %
6	(Ages 35 - 44)	29.09 %	10.00 %	0.00 %
7	(Ages 45 - 54)	34.03 %	0.00 %	0.00 %
8	(Ages 55 - 64)	19.68 %	0.00 %	0.00 %
9	(Ages 65 - 74)	6.86 %	0.00 %	37.50 %
10	(Ages 75 - 84)	2.80 %	0.00 %	62.50 %
Total		100.00 %	100.00 %	100.00 %

older ages, respectively, could be used when securitizing life insurances that are sold to younger people[13] and life settlements that typically involve individuals who are aged above 65.

Each hypothetical mortality bond is further divided into five tranches (I to V) with different attachment and exhaustion points (i.e., $[L, U]$), namely $[1.02, 1.04]$, $[1.04, 1.06]$, $[1.06, 1.08]$, $[1.08, 1.10]$ and $[1.10, 1.12]$. Except parameters w_x, L and U, the hypothetical mortality bonds are identical to the Atlas IX Capital Class B note.

Our goal is to estimate the premium spread for each tranche of the three hypothetical bonds. This goal is accomplished by first solving the following equations for the coupon rate c:

$$E^{\mathbb{Q}_0}\left(\sum_{t=0}^{4}\sum_{s=0}^{3} D_{t+\frac{s+1}{4}} C_{2014+t+\frac{s}{4}} + D_{5+\frac{1}{4}}(C_{2019}+P_{2019})\right) = 1 \quad \text{and} \quad C_{t+\frac{s}{4}} = cP_t,$$

where $\mathbb{Q}_0$ is the canonical measure identified in Section 2.6.3 and the values of $P_{2014}, \ldots, P_{2019}$ in each state of nature are calculated using equations (2.8) to (2.12), on the basis of the sample path of future mortality rates associated with that particular state of nature and the applicable values of w_x, L and U. The premium spread is then estimated by subtract-

[13]Life insurance portfolios that are concentrated on ages below 44 are not uncommon. For instance, in financial year 2012, 74% of the customers who purchased term-life insurances from Lifenet Insurance were aged below 40. Source: http://ir.lifenet-seimei.co.jp/.

Table 2.7: The estimated premium spreads for the hypothetical mortality bonds.

	Tranche I	Tranche II	Tranche III	Tranche IV	Tranche V
Attachment point (%)	102.00	104.00	106.00	108.00	110.00
Exhaustion point (%)	104.00	106.00	108.00	110.00	112.00
	Estimated premium spread (basis points)				
Hypothetical Bond 1					
Model J0	325.00	217.30	192.76	182.70	176.59
Model J1	325.00	266.67	210.63	178.94	166.62
Model J2	325.00	274.19	220.29	186.86	171.94
	Estimated premium spread (basis points)				
Hypothetical Bond 2					
Model J0	546.08	434.86	340.38	280.48	248.94
Model J1	752.94	649.96	558.60	485.48	432.36
Model J2	755.98	658.01	573.23	506.80	459.50
	Estimated premium spread (basis points)				
Hypothetical Bond 3					
Model J0	991.94	530.33	284.89	199.76	175.27
Model J1	489.50	337.69	233.04	184.50	167.30
Model J2	481.44	333.54	231.56	183.97	167.14

ing the rate that collateral account earns (the 3-month LIBOR minus 6 basis points) from the calculated coupon rate.

The estimated premium spreads for all tranches of the three hypothetical bonds are displayed in Table 2.7. We acknowledge that Model J2 is not the most preferred model for the U.S. population, but the results from this model variant are still presented for completeness and for the readers to gauge how the permission of a random age pattern may affect the premium spread estimates. As expected, the estimated premium spreads decrease with the attachment and exhaustion points, because investors should be rewarded less as the bond's riskiness reduces. More noteworthy is that compared to Model J0, Models J1 and J2 (which model the age pattern of mortality jumps with a distinct collection of parameters) yield significantly different premium spread estimates. The differences range from minus 51 to plus 85 percentage points, depending on the bond's attachment and exhaustion points as well as the age weights of the index to which the bond is linked.

Let us first focus on the pricing results for Bond 1. For Tranche I of this bond, all three model variants yield the same premium spread estimate of 325 basis points. This result is not surprising, because this tranche is identical to the Atlas IX Capital Class B note

on which the martingale constraint for the canonical measure is based. Relative to the premium spreads estimated from Model J0, the premium spreads estimated from Models J1 and J2 are larger for Tranches II and III but smaller for Tranche V. This result cannot be explained simply by comparing the jump severity and volatility parameters among the three model variants. However, we may understand this result by examining how the jump severity and volatility parameters interact with the parameters representing the age pattern of mortality jumps. In Figure 2.4 we show the patterns of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$, for $\lambda = 0, 0.5, 1, 2$, implied by Models J0, J1 and J2. This quantity can be understood as the impact of a mortality jump with a magnitude of $\mu_J + \lambda\sigma_J$ on age group x. We now examine the values of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$ for age group 45-54, the age group on which Bond 1's mortality index is most heavily weighted. For $\lambda = 0$ (i.e., for a mortality jump with a moderate magnitude), the value of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$ for age group 45-54 implied by Model J0 is lower than those implied by the other two model variants, but for higher values of λ (i.e., for a mortality jump with a larger magnitude), the opposite is true. This observation offers an explanation as to why Models J0 implies lower premium spreads for the riskier tranches (which have lower attachment and exhaustion points) but higher premium spreads for the less risky tranches. Using a similar argument, we can also explain why Model J2 implies higher premium spreads for Tranches II to V in comparison to Model J1.

We then turn to the pricing results for Bond 2, whose index is most heavily weighted on age groups 15-24 and 25-34. For these two age groups, Model J0 yields the smallest values of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$ for all values of λ we consider. This observation may explain why Model J0 implies the smallest premium spread for all five tranches of Bond 2. On the other hand, except for $\lambda = 0$, Model J2 gives the highest values of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$ for age groups 15-24 and 25-34, accounting for why it produces the highest premium spreads.

Finally, we consider the pricing results for Bond 3, whose index is most heavily weighted on the oldest age group. We have demonstrated in Section 2.5 that Model J0 tends to overestimate the effect of mortality jumps on individuals at older ages. The impact of such an overestimation does have a significant impact in terms of pricing, as indicated by the premium spread estimates shown in Table 2.7. The reason that Model J0 implies the highest premium spread estimates for all tranches of Bond 3 may also be explained by Figure 2.4, from which we observe that Model J0 consistently produces the highest values of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$ for the last age group.

2.6.5 The Effect of Parameter Uncertainty

We have pointed out in Section 2.5 that some of the parameter estimates have rather large standard errors. It is therefore worthwhile to investigate how parameter uncertainty

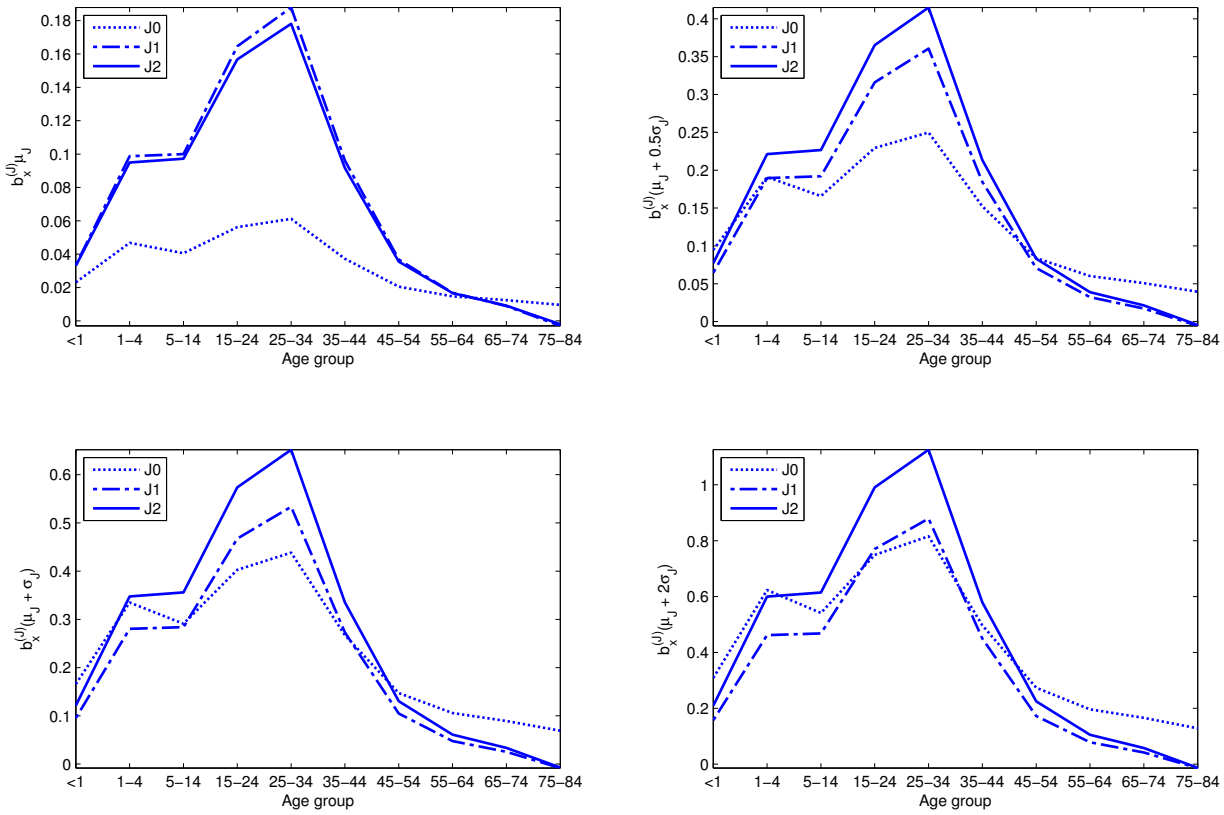


Figure 2.4: The patterns of $b_x^{(J)}(\mu_J + \lambda\sigma_J)$, for $\lambda = 0, 0.5, 1, 2$, implied by Models J0, J1 and J2. Note that for Model J0, $b_x^{(J)} = b_x$.

may affect the risk spread estimates. To accomplish this goal, the following procedure is implemented:

1. Use the parametric bootstrap documented in Appendix A.2 to obtain $M = 5000$ different collections of parameter estimates.

2. For each collection of parameter estimates, simulate $N = 20000$ mortality scenarios. Given these 20000 mortality scenarios, derive the canonical measure and use it to obtain premium spread estimates for all tranches of the three hypothetical bonds under consideration. This step yields, for each tranche of each hypothetical bond, an empirical distribution of $M = 5000$ premium spread estimates.

3. On the basis of the 2.5th and 97.5th percentiles of the empirical distributions derived from Step 2, construct 95% confidence intervals for the premium spread estimates.

These confidence intervals reflect the uncertainty surrounding the premium spread estimates that arises from parameter uncertainty.

The estimated confidence intervals are displayed in Table 2.8. Excluding Tranche I of Bond 1 (which is used to formulate the martingale constraint), the impact of parameter uncertainty on the premium spread estimates for Bonds 1 and 3 is moderate. For these two bonds, the widths of the estimated confidence intervals generated from the three model variants are quite similar, ranging from 6% to 24% of the corresponding central estimates.

Nevertheless, the impact of parameter uncertainty on the premium spread estimates for Bond 2 is far more significant, as indicated by the estimated confidence intervals with widths ranging from 47% to 89% of the corresponding central estimates. This phenomenon may be attributed to the fact that the index of Bond 2 is very heavily weighted on the two age groups (15-24 and 25-34) which are affected the most by mortality jumps. The pricing results for this bond are therefore more sensitive to changes in the jump-effect-related parameters, which are the most uncertain. It can also be seen that for Bond 2, the premium spread estimates derived from Models J1 and J2 are subject to more uncertainty than those from Model J0. This outcome can be explained by the fact that for age groups 15-24 and 25-34, the standard errors of the estimates of $b_x^{(J)}$ in Models J1 and J2 are larger than those of the estimates of b_x in Models J0.

Admittedly, for Bond 2, the confidence intervals derived from our proposed models are significantly wider than those generated from the existing Model J0. Given that the age patterns of mortality jumps do vary and that the payoff from a catastrophic mortality bond does depend heavily on the age patterns of future mortality jumps, the wider confidence intervals may simply be a more realistic reflection of how little we know about the properties of future mortality jumps and hence the true values of catastrophic mortality bonds. The lack of data to reveal precisely the patterns of jumps is definitely a huge limitation of this pricing study, so readers are warned not to overlook the degrees of uncertainty conveyed in the confidence intervals. Business decisions should be made on the basis of the range of possible values, rather than just the central estimate itself.

2.7 Concluding Remarks

In this chapter, we have investigated how the age pattern of transitory mortality jumps can be modeled explicitly. Two new variants of the Lee-Carter model have been proposed. The first variant uses a distinct collection of parameters to capture the age pattern of mortality jumps, while the second variant is a further generalization that allows the age pattern

Table 2.8: 95% interval estimates of the premium spreads for the hypothetical mortality bonds, taken into account of parameter uncertainty.

	Tranche I	Tranche II	Tranche III	Tranche IV	Tranche V
Attachment point (%)	102.00	104.00	106.00	108.00	110.00
Exhaustion point (%)	104.00	106.00	108.00	110.00	112.00
	95% interval estimate of the premium spread (basis points)				
Hypothetical Bond 1					
Model J0	(325.00, 325.00)	(192.70, 241.02)	(170.64, 215.57)	(164.80, 202.92)	(161.94, 195.90)
Model J1	(325.00, 325.00)	(244.30, 286.40)	(191.07, 231.69)	(168.87, 192.61)	(162.44, 174.65)
Model J2	(325.00, 325.00)	(249.04, 294.70)	(196.23, 245.42)	(171.55, 206.47)	(163.59, 185.79)
	95% interval estimate of the premium spread (basis points)				
Hypothetical Bond 2					
Model J0	(389.06, 704.88)	(314.24, 555.54)	(253.12, 433.30)	(208.20, 349.17)	(188.26, 304.46)
Model J1	(565.61, 998.05)	(463.48, 898.37)	(378.12, 787.48)	(304.51, 708.40)	(262.09, 636.79)
Model J2	(567.57, 996.00)	(464.65, 897.57)	(378.73, 812.67)	(318.72, 738.93)	(276.28, 686.54)
	95% interval estimate of the premium spread (basis points)				
Hypothetical Bond 3					
Model J0	(894.25, 1074.5)	(481.77, 571.62)	(257.76, 308.97)	(183.33, 220.99)	(164.98, 193.89)
Model J1	(436.83, 547.00)	(300.37, 368.99)	(209.32, 256.87)	(173.12, 197.69)	(163.33, 173.30)
Model J2	(421.49, 537.58)	(293.04, 366.61)	(206.95, 256.88)	(172.48, 198.88)	(163.21, 173.89)

of future mortality jumps to vary randomly. When applied to historical mortality data from the U.S. and English and Welsh populations, the proposed model variants provide a significantly better fit in comparison to the more restrictive model in which the same age response parameters are used for modeling both general and extreme changes in mortality. We have also found that the way in which the age pattern of mortality jumps is modeled has a significant impact on the pricing of catastrophic mortality bonds.

Our proposed model variants have several pros and cons relative to the model introduced by Cox et al. (2010). While the other model also explicitly captures the age pattern of mortality jumps, our proposed model variants involve less subjectivity, because their parameters can be estimated entirely from historical data. Another advantage of our modeling approach is that it permits a random age pattern of mortality jump effects. On the other hand, the other model is advantageous of having a less stringent data requirement and the ability to incorporate expert judgements. In practice, the analysis of a catastrophic mortality bond is often based on multiple models.[14] We believe that issuers and credit

[14]For instance, according to the S&P Rating Report (www.standardandpoors.com/ratings/ils/en/us),

rating agencies would benefit by making reference to the projection results generated from both our proposed model variants and the model of Cox et al. (2010).

We understand that readers may be concerned with the extent of parameter risk in our modeling approach, because there were only a handful of extreme mortality events occurred during the data sample period. To address this concern, we have used a parametric bootstrapping procedure to quantify the uncertainty surrounding each of the model parameters. It is found that the standard errors of some jump-effect-related parameters (e.g., μ_J and σ_J) are quite large. However, this problem applies to not only our proposed model variants (Models J1 and J2) but also the existing, more restrictive model (Model J0). We have also utilized the parametric bootstrap to analyze the impact of parameter uncertainty on pricing. It is found that the impact depends quite heavily on the age range with which the security being priced is associated. In our illustrations, the impact of parameter risk is only moderate on two out of the three bonds under consideration, but is severe on the remaining bond which is heavily related to the age groups that are subject to the most mortality jump effects.

One may be interested in generalizing the model variants proposed in this chapter to a multi-population version, which would be useful for pricing and analyzing catastrophic mortality bonds that are linked to a group of populations. An important aspect in developing multi-population mortality models is coherence, a property that ensures the mortality projections of related populations do not diverge indefinitely over time (see Li and Lee, 2005; Cairns et al., 2011b). To make the multi-population versions of Models J1 and J2 coherent, we require all populations to share the same set of b_x parameters and the difference between the k_t parameters of any two populations to be mean-reverting over time. Nevertheless, coherence can still be maintained even if the populations being modeled have different jump frequencies, severities and/or age patterns (i.e., parameters p, μ_J and σ_J), as short-term mortality jump effects do not affect long-term mortality trends.

To facilitate exposition, we have restricted our discussion to short-term mortality jumps only. One may, however, use a similar modeling approach to capture intermittent changes in long-term mortality trends. Such an extension is supported by the empirical findings of various researchers, including Sweeting (2011) who noted that there have been multiple changes in the secular mortality trend for English and Welsh males since 1841. The age patterns of intermittent changes in long-term mortality trends merit special attention, because they may reflect important information concerning demographic transitions. For example, the trend changes during the post-war period are believed to be associated with shifts in mortality improvement from younger to older ages (Oeppen and Vaupel, 2002).

the credit rating of the Atlas IX Capital Class B note was derived from five different models provided by Risk Management Solutions.

They may also have pricing implications on various longevity risk transfer products, including pension buy-ins and buy-outs, which have steadily gained popularity over the past decade.

Measuring hedge effectiveness is another possible avenue for future research. By adapting the method of Cairns et al. (2014), we may be able to quantify the amount of extreme mortality risk that can be mitigated by trading a catastrophic mortality bond. Furthermore, by implementing the adapted method with Models J0, J1 and J2, we may be able to seek answers to several interesting research questions. For instance, we may predict by how much hedge effectiveness may be over- or under-estimated if the age pattern of mortality jumps is not explicitly taken into account. We may also better understand how the randomness associated with the age pattern of mortality jumps may affect hedge effectiveness.

Chapter 3

Modeling and Managing Longevity Drift Risk

3.1 Introduction

Depending on the context and intention, longevity risk can be defined in different ways. From a statistical viewpoint, Coughlan et al. (2013) provided the following concise yet complete definition: *"It is a combination of (i) uncertainty surrounding the trend increases in life expectancy and (ii) variations around this uncertain trend that is the real problem. This is what is meant by longevity risk and it arises as a result of unanticipated changes in mortality rates."* Despite the risk encompasses two components, most existing stochastic mortality models capture only the latter. For instance, in the Lee-Carter model (Lee and Carter, 1992), the evolution of mortality over time is typically captured by an autoregressive integrated moving average process, with the special case – random walk with drift – being used the most often. What this modeling method captures is 'diffusion risk', which arises from the variations around a fixed drift that determines the expected trend, but not the uncertainty associated with the trend itself. The collection of models (Models M1-M8) considered by Cairns et al. (2009, 2011a) and Dowd et al. (2010a,b) are subject to the same limitation.

Although often left unmodeled, the risk associated with the trend in mortality improvement does exist. There is profound empirical evidence for trend changes in mortality, exemplified by the findings of Gallop (2006) concerning the mortality improvement in the United Kingdom over the past century. It was found that the average rate of mortality improvement for British males was gently fluctuating around 0.7% per annum over the

period of 1930-32 to 1970-72, after which the rate increased substantially, reaching 2.0% per annum over the period of 1990-92 to 2001-03. Similar trend changes are also observed in other developed countries (see, e.g., Kannisto et al., 1994; Vaupel, 1997). A number of researchers have further confirmed the trend changes in mortality by rigorous statistical tests, including Li et al. (2011) who considered Zivot and Andrew's test, Ahmadi and Li (2014) who used a non-parametric change-point test, and O'Hare and Li (2014) who utilized the CUSUM test that is based on cumulative sums of standardized residuals.

Researchers have also developed mortality models that incorporate trend changes in the past (Li et al., 2011; Ahmadi and Li, 2014, O'Hare and Li, 2014; Coelho and Nunes, 2011; van Berkum et al., 2014; Renshaw and Haberman, 2003), but these models at best can only reflect how historical trend changes may affect the best estimate forecast. They do not capture trend changes as a risk, as they do not allow historical trend changes to recur at random future time points with random extents. For example, in the broken-trend stationary model proposed by Li et al. (2011), it is assumed that the future trend in mortality improvement is always the same as the historical trend after the detected structural break point. The model proposed recently by van Berkum et al. (2014) is less stringent in the sense that multiple trend changes are permitted during the data sample period, but still the drift is assumed to be fixed beyond the last estimated structural break point. An alternative way that has been used to deal with trend changes is to optimize the calibration window, so that the model is fitted to the period of time during which the trend is the most linear (Booth et al., 2002; Denuit and Goderniaux, 2005; Li et al., 2015a). However, excluding a portion of the data does not address the random nature of trend changes.

On the modeling front, we may address the risk associated with trend changes by permitting the drift term(s) in the assumed mortality model to be stochastic. A previous attempt to introduce a stochastic drift was made by Milidonis et al. (2011), who modeled the time-varying factor in the Lee-Carter model using a regime-switching log-normal process with two regimes. The drift term in each regime is permitted to be different, so that the drift of the process may vary as the system switches between the two regimes under an assumed Markov chain. However, as indicated in the estimation results of Milidonis et al. (2011), it is the volatility rather than the drift that separates the two regimes. Therefore, the regime-switching process may be suitable for capturing short-term catastrophic mortality events which are generally accompanied with high mortality volatility, but may not be adequate for capturing the risk arising from drift changes. Another attempt was made by Sweeting (2011), who considered a piece-wise linear regression. Although the slope of the regression line is permitted to change in the future, the probability and extent of future slope changes are calculated in an ad hoc manner. In particular, the probability

is taken as the ratio of the number of observed break points to the total number of data points, while the extent is estimated using the root mean square of the annual changes in the underlying dynamic factor. Alternative mortality models which more explicitly and rigorously address the risk associated with trend changes are yet to be developed.

The first objective of this chapter is to develop a stochastic mortality model that permits the user to quantify not only 'diffusion risk' but also 'drift risk'. Our proposed model is based on the original Cairns-Blake-Dowd (CBD) model (Cairns et al., 2006), under which the dynamics of mortality are driven by two time-varying factors. As usual, the evolution of the time-varying factors is modeled by a bivariate random walk with drift, but on top of that, we permit the drifts themselves to follow another bivariate random walk. This modeling method is justified by the results of Nyblom and Mäkeläinen's (1983) test for random walk coefficients. We call our proposed model the locally-linear CBD model (thereafter the LLCBD model), because the drifts governing the linear increments in the two CBD time-varying factors are different in different time steps. To enable estimation, we first formulate our proposed model in a state-space representation, just as how Mavros et al. (2014), Hári et al. (2008), Pedroza (2006), de Jong and Tickle (2006) and Carter (1996) specified the models they considered. We then use the Kalman filtering technique (Kalman, 1960; Kalman and Bucy, 1961) to estimate the unknown model parameters and retrieve the hidden states (i.e., the two CBD time-varying factors and the two varying drifts) in a recursive manner. The method we use can estimate all parameters in the proposed model in one single stage.

Although the extension of the Lee-Carter model proposed by Hári et al. (2008) possesses a time-varying drift, it is different from our proposed model in various aspects. In terms of objectives, the extension of Hári et al. (2008) was not designed with a motivation to quantify trend risk, and possibly for this reason, it assumes that the drift would fluctuate around its long-term mean. In contrast, with a goal of assessing drift risk, we postulate the dynamics of the drift vector as a random walk. The use of a random walk is in part because of the support from the random walk coefficient test and in part because we have no a priori knowledge about the value and more importantly the existence of the mean of the drift vector. In terms of model structure, the extension of Hári et al. (2008) is built on a reformulation introduced by Girosi and King (2005) whereby the vector of log central death rates are structured to follow a multidimensional random walk, whose drifts are driven by a vector of latent factors which follows a stationary multivariate autoregressive moving average process. To maximize comparability with the original CBD model, our state-space model requires no reformulation and has an observation equation that preserves the parameterization of the original CBD model. Compared to the extension of Hári et al. (2008) in which the latent factors are not straightforward to interpret, our proposed

model contains richer demographical intuitions as all four hidden states in it have their own physical meanings. In spite of the mentioned differences, the estimation results for both models indicate one thing in common: the permission of time-varying drift(s) results in mortality projections that are more robust with respect to changes in the data sample period. Both models can therefore ameliorate the well-known problem that the usual estimator of the drift(s) of a random walk is highly sensitive to (and indeed completely dependent on) the first and last observations of the data series (see, e.g., Li and Chan, 2005; Zhou and Li, 2013).

Having addressed drift risk in modeling, our second objective is to develop a method to manage it. To our knowledge, Cairns (2013) was the first to study how an index-based longevity hedge may be constructed to reduce the hedger's exposure to drift risk. He extended the existing 'delta' hedging method (Cairns, 2011; Luciano et al., 2012; Zhou and Li, 2014) to the 'delta-nuga' hedging method, in which the sensitivities of the liability being hedged and the portfolio of hedging instruments to changes in drifts are matched. While the delta-nuga hedging method has some appeals, it is subject to a few limitations. First, in deriving the delta-nuga hedging strategies, it is assumed that the future values of the time-varying factors in the underlying model are related to the current values in a deterministically linear manner. Hence, if the linear relation does not hold, the hedging results would be sub-optimal. Second, when applied to the CBD/LLCBD model, the delta-nuga hedging method requires exactly four hedging instruments. This stringent requirement may render the method impractical in the early stages of market development when the market does not provide the required number of standardized hedging instruments. Third, as we are going to demonstrate in Section 3.5.5, the delta-nuga hedging method is subject to the singularity problem, which means that the solution to the hedging strategies does not exist when certain combinations of hedging instruments are used. To mitigate these limitations, we propose in this chapter the generalized state-space hedging method.

In the generalized state-space hedging method, we waive the linearity assumption by considering the sensitivities of the liability being hedged and the portfolio of hedging instruments to all future hidden states that are relevant. Similar to the work of Cairns et al. (2014), we derive the hedging strategies by variance minimization. We regard our proposed method as a generalization, because it degenerates to the traditional delta and delta-nuga hedging methods when all future hidden states are deterministically linearly related to the current ones. Based on the sensitivities to all relevant hidden states, the proposed method may also be seen as a complement to several existing static hedging methods, including the methods of q-duration (Coughlan, 2009) and key q-duration (Li and Luo, 2012; Li and Hardy, 2011) that are based on the sensitivities to the death probabilities at selected ages and the method of key K-duration (Tan et al., 2014) that is based on the sensitivities

to the CBD mortality indexes at selected time points. In comparison to the traditional delta and delta-nuga hedging methods, the proposed method is more flexible in terms of the number and type of hedging instruments. Our empirical results indicate that when population basis risk is assumed to be absent, the proposed hedging method can lead to a greater than 85% hedge effectiveness even if only one hedging instrument is used.

The rest of this chapter is organized as follows. In Section 3.2 we explore statistical evidence for stochastic drifts. In Section 3.3 we detail the specification of the proposed LLCBD model. In the same section we also analyze the performance of the proposed model by estimating it to some real mortality data. In Section 3.4 we discuss several additional issues about modeling. In Section 3.5 we present the generalized state-space hedging method and compare it with the traditional delta and delta-nuga hedging methods. In Section 3.6 we illustrate the proposed hedging method with a hypothetical example. Finally, in Section 3.7 we conclude with some suggestions for future research.

3.2 Evidence for Stochastic Drifts

Let $q_{x,t}$ be the *crude* probability that an individual dies between time $t-1$ and t (during year t), given that he/she has survived to age x at time $t-1$. We calculate $q_{x,t}$ by the following approximation:

$$q_{x,t} \approx \frac{1}{1+0.5m_{x,t}},$$

where $m_{x,t}$ is the crude central rate of death at age x and in year t.[1] The simplest version of the CBD model can be expressed as

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \epsilon_{x,t},$$

where $\bar{x}$ represents the average over the sample age range to which the model is fitted, $\epsilon_{x,t}$'s are the sampling errors, which are assumed to be i.i.d. normally distributed with a zero mean and a constant variance of σ_ϵ^2, and $\kappa_1(t)$ and $\kappa_2(t)$ are time-varying stochastic factors, of which the dynamics are assumed to follow a bivariate random walk with constant drifts C_1 and C_2.

We estimate the above model to the mortality data from Canadian male population over a sample age range of 50 to 89 (40 ages) and a sample period of 1941 to 2010 (70

[1]The approximation is exact if deaths are uniformly distributed between two consecutive integer ages.

years), using the method of least squares; that is, the time-varying stochastic factors are obtained using the following objective function:

$$\min_{\kappa_1(t),\kappa_2(t)} \sum_{x=50}^{89} \left(\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) - \kappa_1(t) - \kappa_2(t)(x-\bar{x}) \right)^2, \quad t = 1941, \ldots, 2010.$$

The data are obtained from the Human Mortality Database (2015). Let Δ be the difference operator. If the assumption of constant drifts hold, then the estimated values of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ should fluctuate around their respective sample means $\hat{C}_1$ and $\hat{C}_2$. However, as shown in upper panels of Figure 3.1, the estimates of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ do not seem to follow the expected pattern.

The lower panels of Figure 3.1 depict the 5- and 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$. These moving averages may be considered as proxies for the drifts at different time points. It is clear that the moving averages are time-varying and exhibit random patterns. The observations we made from Figure 3.1 lead to the question as to whether the drifts themselves are stochastic.

We further investigate the stochastic nature of the drift terms by applying a statistical test for random walk coefficients. The following description focuses on C_1, but the test for C_2 can be conducted in a similar manner. Suppose that $\kappa_1(t)$ follows a random walk, with a drift $C_1(t)$ that also follows a random walk itself. The dynamics of $\kappa_1(t)$ can be expressed by the following system of equations:

$$\begin{cases} \Delta\kappa_1(t) &= C_1(t) + \xi(t), \\ C_1(t) &= C_1(t-1) + \upsilon(t), \qquad t = t_a + 1, t_a + 2, \ldots, t_b, \end{cases} \tag{3.1}$$

where $t_a = 1941$ and $t_b = 2010$ represent the beginning and end points of the calibration window, respectively. It is assumed that $\xi(t)$'s and $\upsilon(t)$'s are i.i.d. normally distributed with a zero mean and constant variances, and that $\xi(t)$ and $\upsilon(t)$ are mutually independent.

We let σ_ξ^2 and σ_υ^2 be the variances of $\xi(t)$ and $\upsilon(t)$, respectively. If $\sigma_\upsilon^2 = 0$, then $C_1(t)$ is constant over time rather than following a random walk. It is thus obvious that our goal is to test the null hypothesis $H_0 : \sigma_\upsilon^2 = 0$ against the alternative hypothesis $H_1 : \sigma_\upsilon^2 > 0$. However, as pointed out by LaMotte and McWhorter (1978), it impossible to evaluate the power of such a test. For this reason, they recommended basing the test on the ratio $\sigma_\upsilon^2/\sigma_\xi^2$, which makes the computation of the test's power possible.

We use the locally most powerful invariant (LMPI) test developed by Nyblom and Mäkeläinen (1983) to test $H_0 : \sigma_\upsilon^2/\sigma_\xi^2 = 0$ against $H_1 : \sigma_\upsilon^2/\sigma_\xi^2 > 0$. It follows from equation

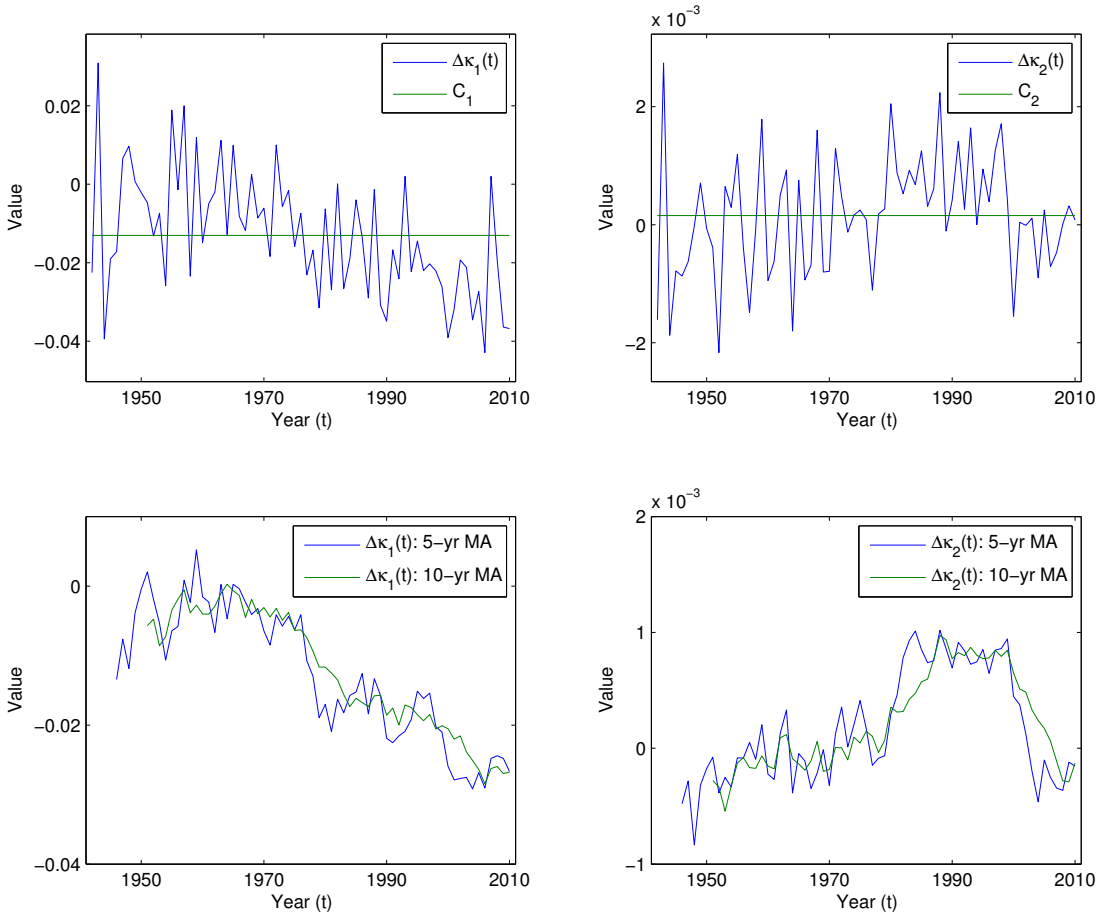


Figure 3.1: The estimated values of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ and their respective means C_1 and C_2 (the upper panels), and the 5-year and 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ (the lower panels).

(3.1) that we can rewrite $\Delta\kappa_1(t)$ as a sum of past innovations:

$$\Delta\kappa_1(t) = C_1(t_a) + \sum_{i=t_a+1}^{t} \upsilon(i) + \xi(t), \qquad t = t_a + 1, t_a + 2, \ldots, t_b.$$

Consequently,

$$\mathrm{cov}(\Delta\kappa_1(s), \Delta\kappa_1(t)) = \sigma_\upsilon^2 \cdot (\min(s,t) - t_a) + \delta_{s,t}\sigma_\xi^2,$$

where $\delta_{s,t}$ equals 1 if $s = t$ and 0 otherwise, and $s, t = t_a + 1, t_a + 2, \ldots, t_b$. It immediately follows that $\Delta\vec{\kappa}_1 = (\Delta\kappa_1(t_a+1), \ldots, \Delta\kappa_1(t_b))' \sim \text{MVN}\left(XC_1(t_a), \sigma_\xi^2(\text{I}_{t_b-t_a} + \sigma_v^2/\sigma_\xi^2 \cdot V)\right)$, where

$$X = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & t_b - t_a \end{pmatrix},$$

MVN represents a multivariate normal distribution, and I_k denotes a k-by-k identity matrix.

Using the generalized Neyman-Pearson lemma, the LMPI test rejects when the LMPI test statistic is greater than some constant that defines the rejection region. On the basis of our set-up, the LMPI test statistic under the null hypothesis can be expressed as

$$\begin{aligned} L_{\text{LMPI}} &= \frac{(\Delta\vec{\kappa}_1 - X\hat{C}_1)' V (\Delta\vec{\kappa}_1 - X\hat{C}_1)}{(\Delta\vec{\kappa}_1 - X\hat{C}_1)'(\Delta\vec{\kappa}_1 - X\hat{C}_1)} \\ &= \frac{\sum_{s,t=t_a+1}^{t_b}\sum^{t_b} (\min(s,t)-t_a)\left(\Delta\kappa_1(s)-\hat{C}_1\right)\left(\Delta\kappa_1(t)-\hat{C}_1\right)}{\sum_{t=t_a+1}^{t_b}\left(\Delta\kappa_1(t)-\hat{C}_1\right)^2} = \frac{\sum_{t=t_a+1}^{t_b}\left(\sum_{s=t}^{t_b}(\Delta\kappa_1(s)-\hat{C}_1)\right)^2}{\sum_{t=t_a+1}^{t_b}\left(\Delta\kappa_1(t)-\hat{C}_1\right)^2}. \end{aligned}$$

The derivation of the distribution of L_{LMPI} is based on the property of invariance in translation. In particular, the denominator of L_{LMPI} can be transformed into a χ^2 variable, while the numerator can be transformed into a linear combination of independent χ^2 variables. Using the results of Nyblom and Mäkeläinen (1983), L_{LMPI} follows the same distribution as

$$\frac{\sum_{k=1}^{t_b-t_a-1} \lambda_{k,t_b-t_a}\left(1 + \lambda_{k,t_b-t_a}\sigma_v^2/\sigma_\xi^2\right) u_k^2}{\sum_{k=1}^{t_b-t_a-1}\left(1 + \lambda_{k,t_b-t_a}\sigma_v^2/\sigma_\xi^2\right) u_k^2},$$

where u_k's are i.i.d. standard normal random variables and $\lambda_{k,t_b-t_a}^{-1} = 2(1-\cos(\pi k/(t_b-t_a)))$ for $k = 1, 2, \ldots, (t_b - t_a - 1)$. At a significance level of α, the rejection region c_α for the LMPI test can be constructed by solving following equation:

$$\alpha = \Pr\left(\frac{L_{\text{LMPI}}}{t_b - t_a - 1} > c_\alpha\right) = \Pr\left(\sum_{k=1}^{t_b-t_a-1}\left(\frac{\lambda_{k,t_b-t_a}}{t_b - t_a - 1} - c_\alpha\right) u_k^2 > 0\right).$$

The value of c_α can be solved numerically by applying Imhof's (1961) method of inversing the characteristic functions. It is found that at $\alpha = 0.05$, the value of $c_{0.05}$ for our tests is 0.4689.

The calculated values of $L_{\text{LMPI}}/(t_b - t_a - 1)$ for testing stochastic drifts in $\kappa_1(t)$ and $\kappa_2(t)$ are 2.0949 and 0.5087, respectively. Because these values are strictly greater than $c_{0.05} = 0.4689$, the null hypotheses for both tests are rejected at a 5% level of significance. The test results recommend modeling both $\kappa_1(t)$ and $\kappa_2(t)$ with stochastic drifts.

3.3 The LLCBD Model

3.3.1 Model Specification

Motivated by the results of the LMPI test, we propose the LLCBD model whereby the drifts in $\kappa_1(t)$ and $\kappa_2(t)$ are stochastic. For notational convenience, we define

$$y_{x,t} := \ln\left(\frac{q_{x,t}}{1 - q_{x,t}}\right).$$

As with the original CBD model, the LLCBD model assumes that

$$y_{x,t} = \kappa_1(t) + \kappa_2(t)(x - \bar{x}) + \epsilon_{x,t}, \quad \text{for} \quad x = x_c, x_c + 1, \ldots, x_d, \tag{3.2}$$

where $[x_c, x_d]$ represents the age range to which the model is being applied, $\bar{x}$ represents the average over the sample age range $[x_a, x_b]$ to which the model is fitted, and $\epsilon_{x,t}$'s are the sampling errors, which are assumed to be i.i.d. normally distributed with a zero mean and a constant variance of σ_ϵ^2. Note that $[x_c, x_d]$ can be wider than the sample age range $[x_a, x_b]$, because the smooth age-interaction function $x - \bar{x}$ in the model permits extrapolation across the age dimension[2]. Note also that the quantity being modeled ($q_{x,t}$) is the *crude* conditional death probability, so the sampling error, $\epsilon_{x,t}$, in equation (3.2) measures the uncertainty due to variation in the actual number of deaths, provided that the unobserved underlying death probability

$$\tilde{q}_{x,t} = \frac{e^{\kappa_1(t) + (x - \bar{x})\kappa_2(t)}}{1 + e^{\kappa_1(t) + (x - \bar{x})\kappa_2(t)}} \tag{3.3}$$

[2]This idea is similar to forecasting into the future. The estimated model is allowed to "forecast" to higher ages.

is known. The variability of $\epsilon_{x,t}$'s depends critically on the size of the population being modeled. Other things equal, the larger the population size is, the smaller the variance of $\epsilon_{x,t}$ (i.e., σ_ϵ^2) is.

The difference between the original CBD and the LLCBD models lies in the way in which the time-varying stochastic factors $\kappa_1(t)$ and $\kappa_2(t)$ are modeled. Specifically, for $i = 1, 2$, $\kappa_i(t)$ in the LLCBD model follows the following system of stochastic processes:

$$\begin{cases} \kappa_i(t) &= C_i(t-1) + \kappa_i(t-1) + \eta_i(t), \\ C_i(t) &= C_i(t-1) + \eta_{i+2}(t). \end{cases}$$

Equivalently speaking, $\kappa_i(t)$ follows a random walk with a stochastic drift, which itself follows another random walk. We describe this extension of the CBD model as 'locally linear', since the drifts in $\kappa_1(t)$ and $\kappa_2(t)$ at different discrete time steps are different. We further assume that the vector of innovations $\vec{\eta}_t = (\eta_1(t), \eta_2(t), \eta_3(t), \eta_4(t))'$ possesses no serial correlation and follows a multivariate normal distribution with a zero mean vector and a covariance matrix Q. Note that $\vec{\eta}_t$ measures the uncertainty surrounding the *unobserved* underlying death probability (i.e., systematic longevity risk). This piece of uncertainty exists even if the number of persons-at-risk is infinitely large.

By design, all four hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model are stochastic. Therefore, Q in the LLCBD model must be positive definite, so that the multivariate normal distribution which $\vec{\eta}_t$ follows is non-degenerate. Let $Q_{i,j}$ be the (i, j)th element in matrix Q. We permit the off-diagonal elements in Q (i.e., $Q_{i,j}$ for $i \neq j$) to be non-zero so that the innovations can be statically correlated with one another.

All four hidden states in the LLCBD model are interpretable. As in the original CBD model, $\kappa_1(t)$ and $\kappa_2(t)$ respectively represent the level and slope of the mortality curve (the curve of $q_{x,t}$ in year t) after a logit transformation. Hence, a reduction in $\kappa_1(t)$ indicates an overall mortality improvement, while an increase in $\kappa_2(t)$ means that mortality (in logit scale) at younger ages (below the mean $\bar{x}$ of the sample age range) improves more rapidly than at older ages. Because $C_1(t)$ and $C_1(t)$ respectively govern the rates of change in $\kappa_1(t)$ and $\kappa_2(t)$, we can interpret $C_1(t)$ to mean the (local) pace of mortality improvement and $C_2(t)$ to mean the (local) change in the age distribution of mortality improvements.

To facilitate estimation and analyses, it is more convenient to express the LLCBD model as a linear Gaussian state-space model comprising of an observation equation and an unobservable state process. We let $\vec{y}_t = (y_{x_c,t}, y_{x_c+1,t}, \ldots, y_{x_d,t})'$ be the vector of observations at time t. The observation equation is given by

$$\vec{y}_t = B\vec{\alpha}_t + \vec{\epsilon}_t,$$

where

$$B = \begin{pmatrix} 1 & x_c - \bar{x} & 0 & 0 \\ 1 & x_c + 1 - \bar{x} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_d - \bar{x} & 0 & 0 \end{pmatrix},$$

$\vec{\alpha}_t = (\kappa_1(t), \kappa_2(t), C_1(t), C_2(t))'$ and $\vec{\epsilon}_t = (\epsilon_{x_c,t}, \epsilon_{x_c+1,t}, \ldots, \epsilon_{x_d,t})'$ respectively represent the vector of unobservable states and the vector of error terms at time t. Given the distributional assumptions we made, $\vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \sigma_\epsilon^2 \cdot \text{I}_{x_d - x_c + 1})$.

The unobservable state process can be expressed as

$$\vec{\alpha}_t = A\vec{\alpha}_{t-1} + \vec{\eta}_t,$$

where

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and $\vec{\eta}_t$, as previously defined, is the time-t innovation vector which follows $\text{MVN}(0, Q)$.

The state-space specification above is quite general and can be adapted easily to yield different model variants. Most notably, we can recover the original CBD model by setting

$$Q = \begin{pmatrix} Q_{1,1} & Q_{1,2} & 0 & 0 \\ Q_{2,1} & Q_{2,2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $Q_{i,j}$ for $i, j = 1, 2$ are free parameters, so that the drift terms are forced to be constant. In this special case, $\vec{\eta}_t$ follows a degenerate multivariate normal distribution. To maintain the stochastic nature of $\kappa_1(t)$ and $\kappa_2(t)$, the sub-matrix

$$Q^* := \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \tag{3.4}$$

must be positive definite.

3.3.2 Estimation

To illustrate, we fit the LLCBD model to the mortality data of Canadian males over a calibration window of 1941-2010 and an age range of 50-89. As a comparison, we also fit the original CBD model to the same data set. The estimation of unknown parameters and retrieval of the hidden states are accomplished by the EM algorithm and the Kalman filter, the details of which are provided in the Appendix B.1. For fair comparison, we use same initial values for both the original CBD model and the LLCBD model. The initial values are selected to be the corresponding estimates of $\kappa_1(t_a), \kappa_2(t_a), C_1(t_a)$ and $C_2(t_a)$ using the estimation method from the paper of Cairns et al. (2006)[3].

Table 3.1 displays the estimated parameters (i.e., σ_ϵ^2 and Q) for both the original CBD model and the LLCBD model. Also shown in the table are the parameters' confidence intervals, which are computed by bootstrapping (Cavanaugh and Shumway, 1997; Stoffer and Wall, 1991). We observe that the permission of stochastic drifts leads to only a minimal change in σ_ϵ^2, but results in rather significant reductions in $Q_{1,1}$ and $Q_{2,2}$. The latter observation is because in the LLCBD model, part of the volatilities of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ is captured by $Q_{3,3}$ and $Q_{4,4}$.

In Figure 3.2 we show the values of the hidden states in both estimated models. The values for years 1941 to 2010 are retrieved from the historical data[4], whereas those for years 2011 and onwards are forecasted. The degree of forecast uncertainty for each hidden state can be seen from the corresponding fan chart, which shows the central 10% prediction interval with the heaviest shading, surrounded by the 20%, 30%, ..., 90% prediction intervals with progressively lighter shadings. The line in the middle of each fan chart represents the corresponding median forecast.

Let us first focus on $C_1(t)$ and $C_2(t)$. Under the LLCBD model, the retrieved values of $C_1(t)$ and $C_2(t)$ vary considerably over the calibration window, providing another piece of evidence against the assumption of constant drifts; the median forecasts of $C_1(t)$ and $C_2(t)$ are in line with the retrieved values in the recent past, and are surrounded by ample forecast uncertainty. In sharp contrast, under the original CBD model, the retrieved and forecasted values of $C_1(t)$ and $C_2(t)$ are constant over time, and are roughly equal to the average values of $C_1(t)$ and $C_2(t)$ in the LLCBD model retrieved over the calibration window; the forecasted values are apparently biased high.

[3]In the paper of Cairns et al. (2006), the authors first transform $q_{x,t}$ into $\text{logit}(q_{x,t})$. Then $\kappa_1(t)$ and $\kappa_2(t)$ for each time t are estimated from the model $\text{logit}(\text{q}_{\text{x,t}}) = \kappa_1(\text{t}) + \kappa_2(\text{t})(\text{x} - \bar{\text{x}}) + \epsilon_{\text{x,t}}$ using least squares.

[4]The retrieved states are the expectations of the hidden states given all the information that we have (as detailed in Appendix B.1).

Table 3.1: The estimated values of σ_ϵ^2 and Q and the corresponding 95% confidence intervals, the original CBD model and the LLCBD model.

Parameter	Estimate	95% Confidence Interval	
The original CBD model			
σ_ϵ^2	2.2462×10^{-3}	$(2.1291 \times 10^{-3},$	$2.3664 \times 10^{-3})$
$Q_{1,1}$	2.0164×10^{-4}	$(1.1644 \times 10^{-4},$	$3.0781 \times 10^{-4})$
$Q_{1,2}$	8.3876×10^{-7}	$(-3.5692 \times 10^{-6},$	$5.0612 \times 10^{-6})$
$Q_{2,2}$	7.0298×10^{-7}	$(3.4742 \times 10^{-7},$	$1.1430 \times 10^{-6})$
The LLCBD model			
σ_ϵ^2	2.2109×10^{-3}	$(2.1237 \times 10^{-3},$	$2.3735 \times 10^{-3})$
$Q_{1,1}$	6.6911×10^{-5}	$(2.4100 \times 10^{-5},$	$1.3649 \times 10^{-4})$
$Q_{1,2}$	3.2010×10^{-6}	$(6.1500 \times 10^{-7},$	$6.1600 \times 10^{-6})$
$Q_{1,3}$	-7.0967×10^{-6}	$(-1.6900 \times 10^{-5},$	$1.3950 \times 10^{-5})$
$Q_{1,4}$	-1.0880×10^{-6}	$(-2.0000 \times 10^{-6},$	$1.6600 \times 10^{-7})$
$Q_{2,2}$	1.7979×10^{-7}	$(4.9750 \times 10^{-8},$	$4.8300 \times 10^{-7})$
$Q_{2,3}$	-1.3259×10^{-7}	$(-7.7000 \times 10^{-7},$	$1.1300 \times 10^{-6})$
$Q_{2,4}$	-6.4419×10^{-8}	$(-1.0600 \times 10^{-7},$	$8.0700 \times 10^{-9})$
$Q_{3,3}$	4.4151×10^{-6}	$(2.0650 \times 10^{-7},$	$1.1300 \times 10^{-5})$
$Q_{3,4}$	-4.3974×10^{-8}	$(-2.7950 \times 10^{-7},$	$2.8450 \times 10^{-7})$
$Q_{4,4}$	2.5681×10^{-8}	$(2.6850 \times 10^{-9},$	$5.5850 \times 10^{-8})$

The patterns of the retrieved values of $C_1(t)$ and $C_2(t)$ in the LLCBD model are informative. The trend in $C_1(t)$ appears to fluctuate around a constant prior to the 1970s, but then reduces rapidly over the next two decades. The rapid reduction in $C_1(t)$ echoes the observations made by Kannisto et al. (1994) and Vaupel (1997) that the rates of mortality improvement in the developed world significantly accelerated in the 1970s. The trend in $C_2(t)$ also seems stable prior to the 1970s, but then the stability ceases. The pattern of $C_2(t)$ suggests that the age distribution of mortality improvements for Canadian males has undergone rapid changes over the past four decades.

Next, we turn to the patterns of $\kappa_1(t)$ and $\kappa_2(t)$ over time. The dynamics of these two hidden states are of our particular interest, because the death probabilities for all ages at time t are determined entirely by the values of $\kappa_1(t)$ and $\kappa_2(t)$. For $\kappa_1(t)$ and $\kappa_2(t)$, the two models yield similar retrieved values, but highly different forecasts. For the LLCBD model, the gradients of the median forecasts and the retrieved values in the recent past are quite consistent with each other, but this consistency does not apply to the original CBD model. These observations are the consequences of the aforementioned differences

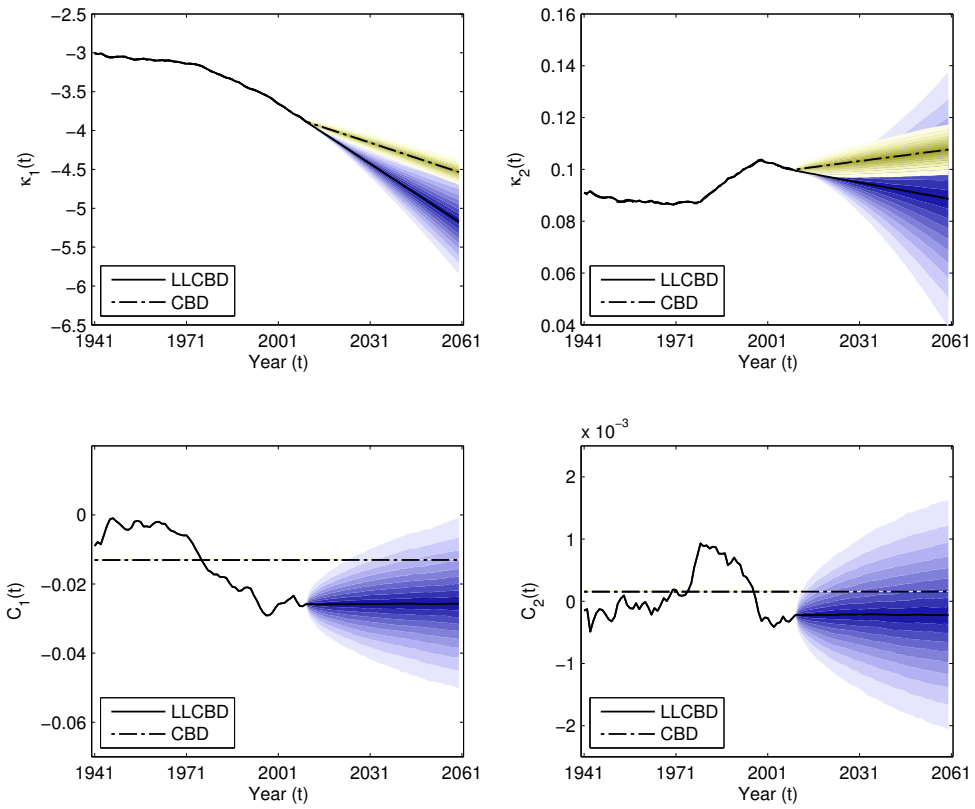


Figure 3.2: The retrieved (1941-2010) and forecasted (2011-2060) values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the original CBD model and the LLCBD model.

in the patterns of $C_1(t)$ and $C_2(t)$ – which determine the expected speed at which $\kappa_1(t)$ and $\kappa_2(t)$ vary – generated from the two models. It is also noteworthy to compare the levels of forecast uncertainty. By permitting stochastic drifts, the LLCBD model results in more conservative prediction intervals for $\kappa_1(t)$ and $\kappa_2(t)$ (and hence for $q_{x,t}$) in the long run. This outcome is not surprising, because the randomness associated with the drifts contributes to the uncertainty surrounding the forecasts of $\kappa_1(t)$ and $\kappa_2(t)$.

Finally, we remark that the hidden states retrieved over 1941-2010 are subject to uncertainty. Figure 3.3 shows the 95% confidence intervals for the retrieved hidden states in the LLCBD model. Following Shumway and Stoffer (2006), the 95% confidence interval for $\kappa_1(t)$ is calculated as $\kappa_1(t) \pm 1.96\sqrt{\text{Var}(\kappa_1(t))}$, and those for the other hidden states are calculated in a similar manner. The variances of the retrieved states are computed using

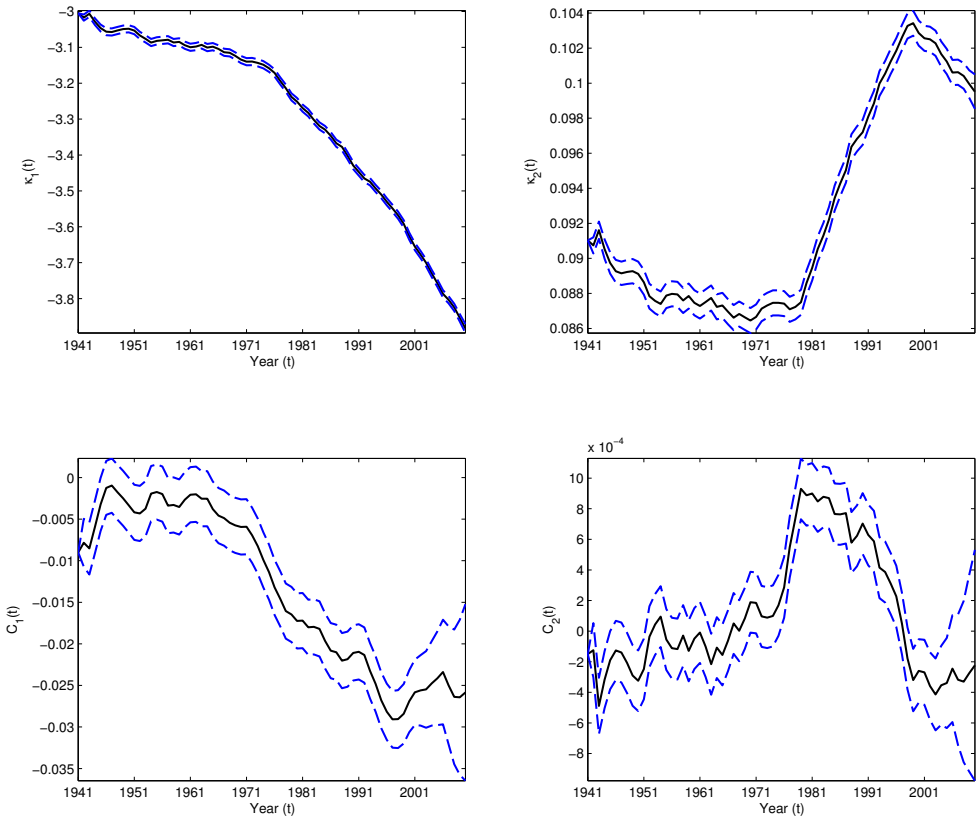


Figure 3.3: The retrieved hidden states (solid lines) in the LLCBD model and their 95% confidence intervals (dashed lines), 1941-2010.

the Kalman filter and Kalman smoother, which are detailed in the Appendix B.1.

3.3.3 Goodness-of-fit

We first evaluate the fit of the LLCBD model to the historical data with a test suggested by Harvey (1990). The test is based on the model's vector of prediction errors, $\vec{w}_t = (w_{x_a,t}, w_{x_a+1,t}, \ldots, w_{x_b,t})'$, which can be computed as

$$\vec{w}_t = \vec{\tilde{y}}_t - BA\mathrm{E}[\vec{\alpha}_{t-1}], \qquad t = t_a + 1, t_a + 2, \ldots, t_b,$$

where $\vec{\tilde{y}}_t$ denotes the realization of $\vec{y}_t$, $\mathrm{E}[\vec{\alpha}_{t-1}]$ represents the retrieved vector of hidden states at time $t-1$, and $BA\mathrm{E}[\vec{\alpha}_{t-1}]$ is the one-step ahead predicted value of $\vec{y}_t$. According

to Harvey (1990), the goodness-of-fit of a state-space model can be evaluated through the prediction error variance σ_p^2 and the prediction error mean deviation $\mathcal{D}$, which can approximated as

$$\sigma_p^2 \approx \frac{1}{(x_b - x_a + 1)(t_b - t_a)} \sum_{x=x_a}^{x_b} \sum_{t=t_a+1}^{t_b} w_{x,t}^2$$

and

$$\mathcal{D} \approx \frac{1}{(x_b - x_a + 1)(t_b - t_a)} \sum_{x=x_a}^{x_d} \sum_{t=t_a+1}^{t_b} |w_{x,t}|,$$

respectively. If the sample size is large and the model is specified correctly, then $w_{x,t}$'s are i.i.d. normal random variables with a zero mean and a constant variance σ_p^2. Under this condition[5], the prediction error mean deviation $\mathcal{D}$ would converge in probability to $(2/\pi)^{0.5}\sigma_p$. Therefore, if the model provides an adequate fit to the historical data, then the value of $(2/\pi)\sigma_p^2/(\mathcal{D}^2)$ should be close to 1. For the fitted LLCBD model, $\sigma_p^2 = 0.0023$, $\mathcal{D} = 0.0370$ and thus $(2/\pi)\sigma_p^2/(\mathcal{D}^2) = 1.07$. The calculated value of $(2/\pi)\sigma_p^2/(\mathcal{D}^2)$ indicates an adequate fit.

Next, we compare the fit of the LLCBD model with that of the original CBD model. When formulated as a Gaussian state-space model, the original CBD model is nested in the LLCBD model. Therefore, we can evaluate the relative goodness-of-fit of the two models by the Akaike information criterion (AIC), which is defined as

$$\text{AIC} = 2(\mathcal{N} - \ln(\hat{\mathcal{L}})),$$

where $\mathcal{N}$ and $\hat{\mathcal{L}}$ represent the number of parameters and the maximized likelihood value, respectively. A model with a smaller AIC value is more preferred. We remark here that in a state-space formulation, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$ are regarded as hidden states rather than model parameters. Hence, for example, the total number of parameters in the LLCBD model is 11, encompassing σ_ϵ^2 and 10 distinct elements of matrix Q.

In Table 3.2 we report the values of $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ and AIC for each model we estimated. The results indicate that the LLCBD model provides a significantly better fit than the original CBD model, taken into account the additional parameters it contains.

[5]If $v_{x,t}$'s are i.i.d. but *not* normal, then $\mathcal{D}$ would converge to another constant instead.

Table 3.2: The values of $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ and AIC for the original CBD model and the LLCBD model.

Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	4465.05	−8922.11	4
LLCBD	4487.89	−8953.78	11

3.3.4 Forecasting Performance

We now perform two tests to evaluate the forecasting performance of the models under consideration. The first test is the 'contracting horizon backtest' previously considered by Dowd et al. (2010b) and Lee and Miller (2001). The test is based on the accuracy of the projections of $\ln(q_{x,t}/1 - q_{x,t})$ in year $t = 2010$, using models that are estimated to data over different calibration windows. In particular, the first forecast is derived from data over 1941-1971, the second forecast is derived from data over 1941-1972, and so on. As the end point of calibration window becomes closer to 2010, the forecasted value of $\ln(q_{x,2010}/1 - q_{x,2010})$ should converge to the actual value. We may regard the forecasting performance of a model as good if the model yields projections that are close to the actual value, no matter what the calibration window is. The result of this test for $x = 60$ is shown in Figure 3.4. Except the first few, the median forecasts produced by the LLCBD model are fairly close to the actual value realized in 2010, but those generated from the original CBD model are consistently biased high. In addition, by comparing the proportions of the 95% prediction intervals that encompass the actual value, we may infer that the LLCBD model provides a more adequate provision of uncertainty. The results for other values of x in the sample age range are similar and are therefore not shown.

In the second test, we estimate the models to restricted calibration windows, and then compare the forecasted values produced by the models with the actual values that are not used in fitting the models. We consider 39 restricted sample periods, ranging from 1941-1971 to 1941-2009. The comparisons are made on the basis of two metrics: Mean Error (ME) and Mean Squared Error (MSE). The results of this test are tabulated in Table 3.3. For instance, the mean error of -0.0675 for the LLCBD model with a calibration window of 1941-1995 is computed by averaging the errors (defined as the actual value less the forecasted value) made in the forecasts of $\ln(q_{x,t}/1 - q_{x,t})$ for $x = 50, \ldots, 89$ and $t = 1996, \ldots, 2010$. It can be seen that on the basis of all three metrics, the LLCBD model consistently yields better forecast accuracy in comparison to the original CBD model.

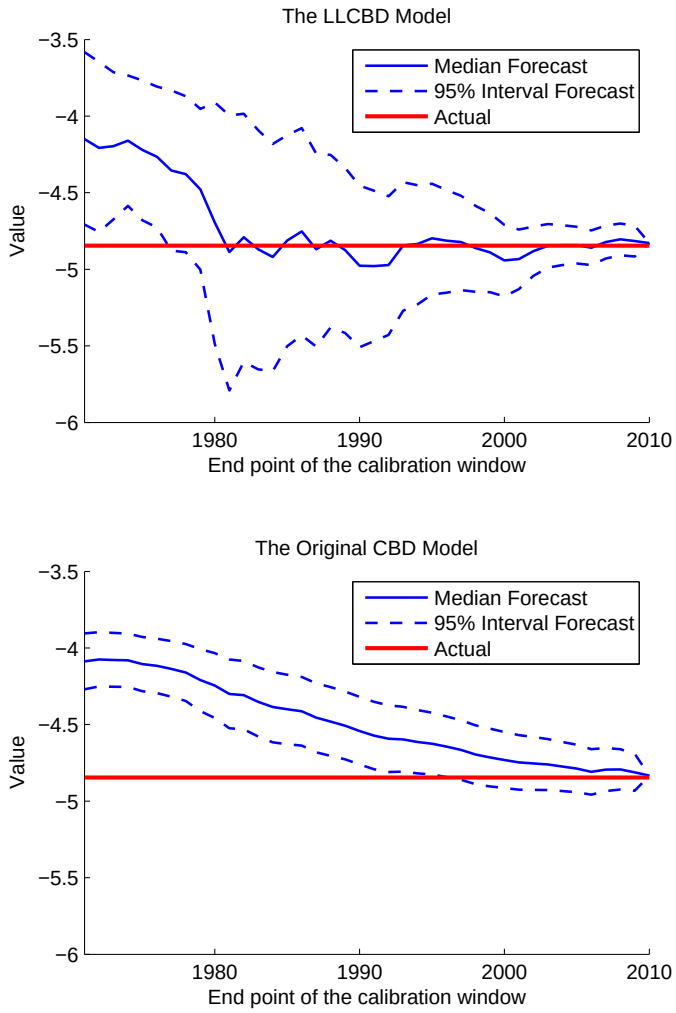


Figure 3.4: The median and 95% interval forecasts of $\ln(q_{60,2010}/1 - q_{60,2010})$, generated from the CBD and LLCBD models that estimated to data over different calibration windows. The starting point of the calibration windows is always 1941 but end points range from 1971 to 2009.

3.3.5 Robustness

When modeling mortality dynamics, it is reasonable to incorporate the most recent data. However, there is no consensus among researchers as to what length of calibration window should be used. It has been demonstrated extensively that mortality forecasts produced by traditional projection models, in which the drift term(s) is/are assumed to be constant, are highly sensitive to the length of the calibration window used. While a longer calibration

Table 3.3: The Mean Errors (ME) and Mean Squared Errors (MSE) for the forecasts of $\ln(q_{x,t}/1-q_{x,t})$ produced by the original CBD model and the LLCBD model, using data over different calibration windows.

	ME		MSE			ME		MSE	
Calibration window	LLCBD	CBD	LLCBD	CBD	Calibration window	LLCBD	CBD	LLCBD	CBD
1941-1971	−0.1956	−0.2267	0.0927	0.1120	1941-1991	−0.0054	−0.1204	0.0195	0.0250
1941-1972	−0.1775	−0.2493	0.0797	0.1209	1941-1992	0.0009	−0.1116	0.0169	0.0225
1941-1973	−0.1925	−0.2563	0.0850	0.1230	1941-1993	−0.0534	−0.1267	0.0155	0.0256
1941-1974	−0.2409	−0.2661	0.1027	0.1262	1941-1994	−0.0528	−0.1215	0.0136	0.0238
1941-1975	−0.2223	−0.2611	0.0871	0.1203	1941-1995	−0.0675	−0.1239	0.0139	0.0239
1941-1976	−0.2025	−0.2623	0.0755	0.1192	1941-1996	−0.0616	−0.1183	0.0131	0.0222
1941-1977	−0.1264	−0.2434	0.0548	0.1119	1941-1997	−0.0638	−0.1163	0.0144	0.0216
1941-1978	−0.1151	−0.2338	0.0521	0.1059	1941-1998	−0.0587	−0.1107	0.0177	0.0210
1941-1979	−0.0567	−0.2046	0.0407	0.0909	1941-1999	−0.0507	−0.1057	0.0187	0.0201
1941-1980	−0.0691	−0.2049	0.0117	0.0826	1941-2000	−0.0095	−0.0842	0.0143	0.0145
1941-1981	−0.0232	−0.1848	0.0112	0.0683	1941-2001	0.0047	−0.0688	0.0107	0.0116
1941-1982	−0.0668	−0.1933	0.0146	0.0686	1941-2002	−0.0089	−0.0669	0.0079	0.0112
1941-1983	−0.0386	−0.1768	0.0135	0.0584	1941-2003	−0.0199	−0.0657	0.0069	0.0109
1941-1984	−0.0213	−0.1649	0.0146	0.0513	1941-2004	−0.0059	−0.0506	0.0055	0.0087
1941-1985	−0.0683	−0.1741	0.0171	0.0512	1941-2005	−0.0020	−0.0411	0.0057	0.0079
1941-1986	−0.0822	−0.1745	0.0159	0.0499	1941-2006	0.0244	−0.0143	0.0061	0.0064
1941-1987	−0.0389	−0.1545	0.0130	0.0409	1941-2007	0.0001	−0.0289	0.0059	0.0071
1941-1988	−0.0835	−0.1644	0.0224	0.0407	1941-2008	−0.0187	−0.0375	0.0062	0.0074
1941-1989	−0.0554	−0.1507	0.0190	0.0353	1941-2009	−0.0171	−0.0293	0.0059	0.0065
1941-1990	−0.0043	−0.1277	0.0176	0.0281					

window permits us to incorporate more information from the historical data, it generally leads to a forecast that is not sufficiently consistent with the recent trend. This problem, as we are about to demonstrate, may be ameliorated by permitting stochastic drifts.

We consider four calibration windows which have the same end point but different starting points: 1941-2010, 1951-2010, 1961-2010 and 1971-2010. We first perform the LMPI test for the four calibration windows (see Table 3.4). For all four calibration windows, the null hypothesis of a constant drift in $\kappa_1(t)$ is rejected at the 5% significant level; and for all but only one calibration window, the null hypothesis of a constant drift in $\kappa_2(t)$ is rejected at the 10% significant level.

Then, we estimate both the original CBD model and the LLCBD model to data over the four calibration windows. For each model, we examine how the resulting forecasts may change as the starting point of the calibration window moves.

Table 3.4: The LMPI test results (test statistic and critical values at 5% and 10% significance levels) for different calibration windows.

	Test statistic		Critical Value	
Calibration window	$\Delta\kappa_1(t)$	$\Delta\kappa_2(t)$	5%	10%
1941-2010	2.0949	0.5087	0.4689	0.3485
1951-2010	2.2383	0.3515	0.4709	0.3503
1961-2010	1.9063	0.2661	0.4719	0.3529
1971-2010	0.9425	0.4444	0.4728	0.3544

Figure 3.5 and 3.6 depict the forecasts of the hidden states in both models on the basis of the four different calibration windows. For the original CBD model, the four calibration windows lead to noticeably different estimates of $C_1(t)$ and $C_2(t)$, and hence considerably different rates of change in $\kappa_1(t)$ and $\kappa_2(t)$. The four median forecasts of $\kappa_1(t)$ and $\kappa_2(t)$ are clearly diverging, while the four fan charts are far from being overlapping one another. Despite the forecasts based on the calibration window starting in 1971 are somewhat consistent with the recent trends, the consistency diminishes significantly as the calibration window begins earlier.

Compared to the original CBD model, the LLCBD model produces median forecasts that are substantially more robust with respect to changes in the beginning point of the calibration window. Regardless of how long the calibration window is, the consistency of the median forecasts with the recent trends remains. These features may be attributed again to the permission of varying drifts, so that the projected rates of change in $\kappa_1(t)$ and $\kappa_2(t)$ are in line with the rates of change in the recent past rather than being close to the average rates of change over the calibration window.

The widths of the fan charts for $C_1(t)$ and $C_2(t)$ in the LLCBD model deserve a few comments. For $C_1(t)$, the longer the calibration window is, the wider the fan chart is. This relationship is expected, as more historical variations are incorporated into the model when the calibration window lengthens. However, the opposite is true for $C_2(t)$. This apparently nonintuitive relationship may be attributed to the fact that volatility of the retrieved values of $C_2(t)$ in the earlier decades is much smaller. As we begin the calibration window earlier, the calibration window covers a longer period of low $C_2(t)$ volatility while the period of high $C_2(t)$ volatility covered remains unchanged, so the (average) $C_2(t)$ volatility captured by estimated model becomes smaller.

Figure 3.7 shows, for both models, the forecasts of $q_{x,t}$ at ages 65 and 75 that are based on the four different calibration windows. In terms of $q_{x,t}$, the difference between the robustness of the two models is even more apparent.

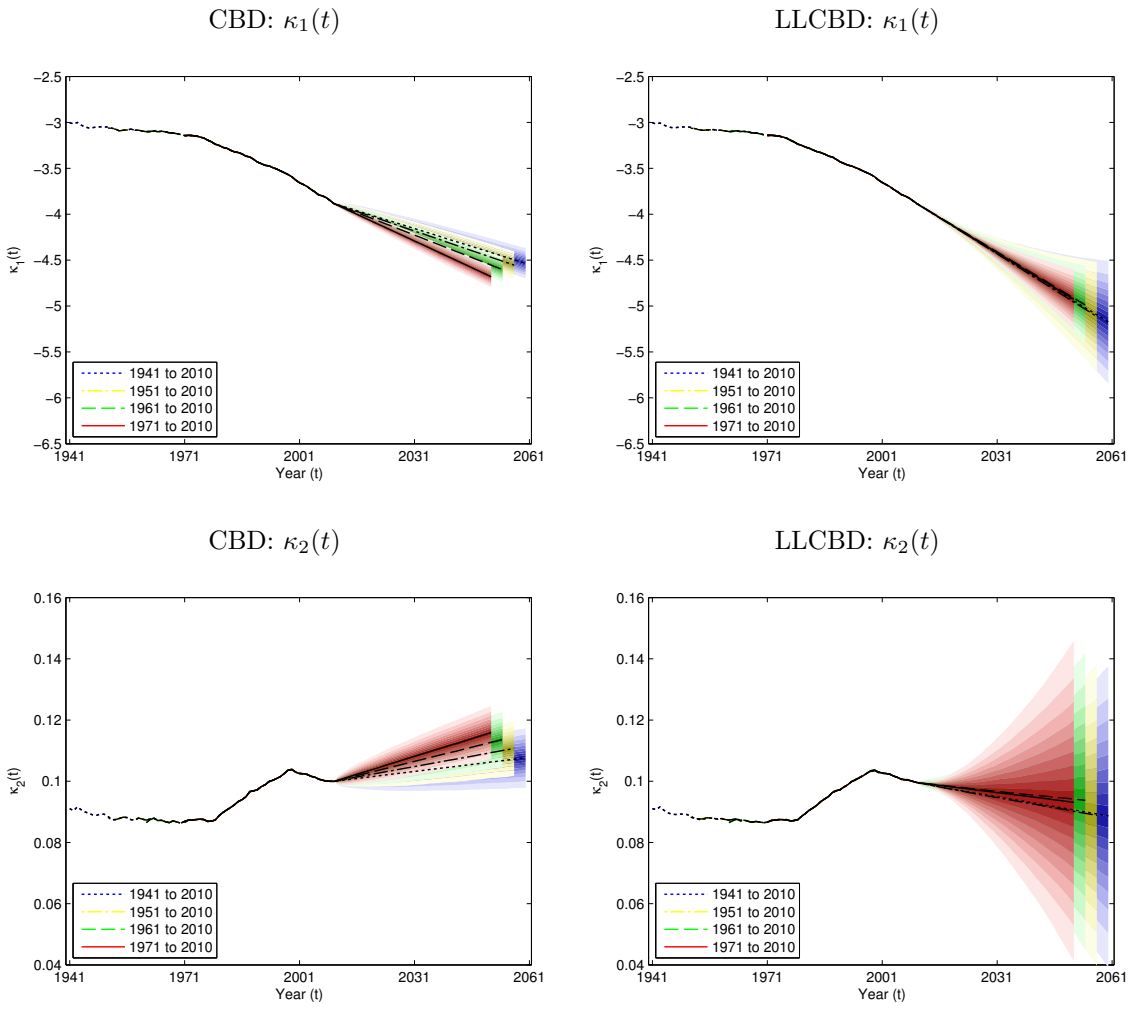


Figure 3.5: Forecasts of the hidden states $\kappa_1(t)$ and $\kappa_2(t)$ in the original CBD model and the LLCBD model that are fitted to data over four calibration windows: 1941-2010, 1951-2010, 1961-2010, 1971-2010.

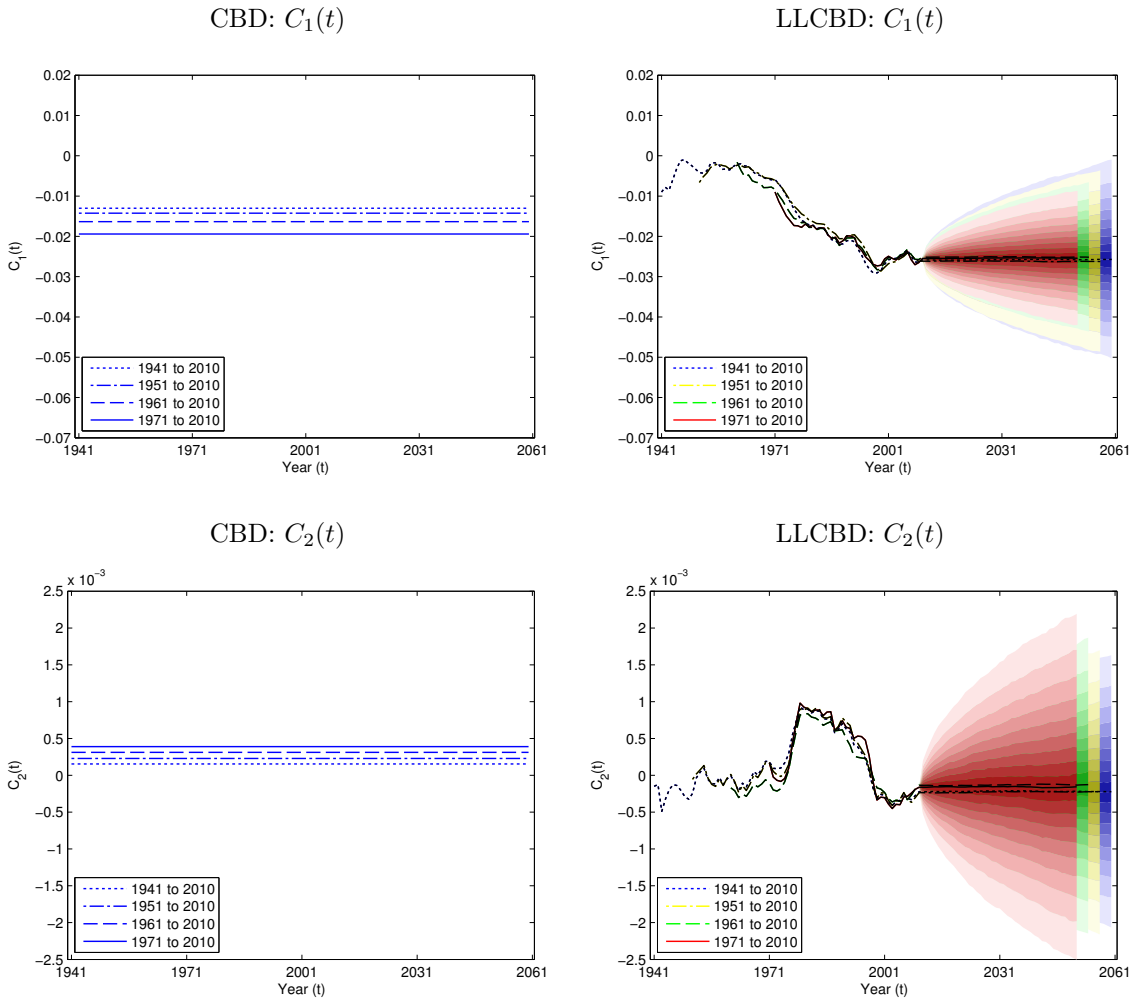


Figure 3.6: Forecasts of the hidden states $C_1(t)$ and $C_2(t)$ in the original CBD model and the LLCBD model that are fitted to data over four calibration windows: 1941-2010, 1951-2010, 1961-2010, 1971-2010.

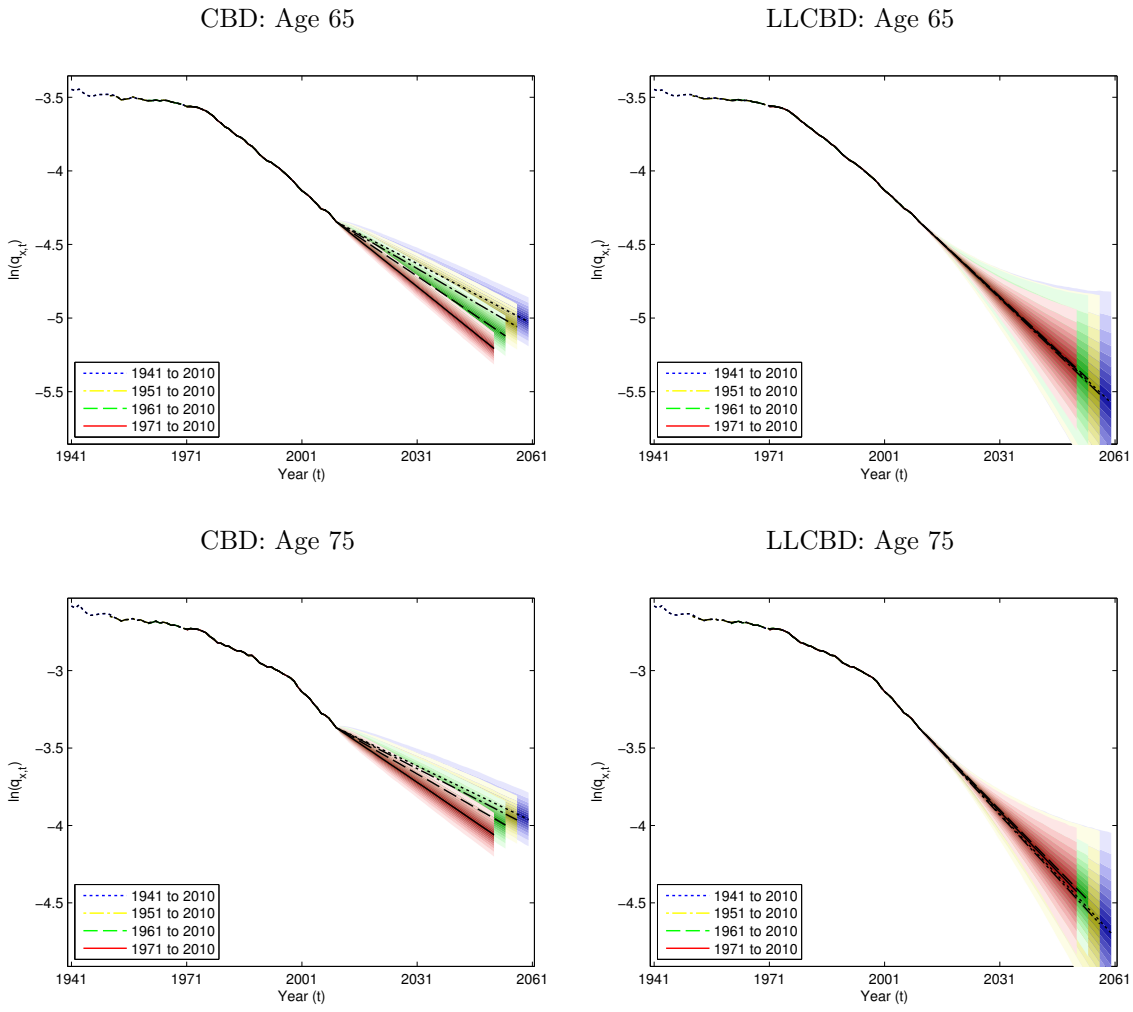


Figure 3.7: Forecasts of $\ln(q_{x,t})$ at $x = 65, 75$ produced by the original CBD model and the LLCBD model that are fitted to data over four calibration windows: 1941-2010, 1951-2010, 1961-2010, 1971-2010.

Still, there exists small variation in the width of the LLCBD fan charts. As $\ln(q_{x,t})$ is a function of $\kappa_1(t)$ and $\kappa_2(t)$, the uncertainty surrounding $\ln(q_{x,t})$ depends on the uncertainty surrounding $C_1(t)$ and $C_2(t)$. In Figure 3.7, the two ages, $x = 65$ and $x = 75$, considered are close to $\bar{x} = 69.5$. When x is close to $\bar{x}$, the coefficient $(x - \bar{x})$ of $\kappa_2(t)$ is small, which means the uncertainty surrounding $C_2(t)$ has a relatively small impact on the uncertainty surrounding $\ln(q_{x,t})$. For this reason, the fan charts of $\ln(q_{x,t})$ have similar patterns to the

fan charts of $C_1(t)$: the longer the calibration window is, the wider the fan chart is.

As in the forecasts of the hidden states, in the *long run* the interval forecasts of $q_{x,t}$ produced by the LLCBD model are more conservative than those generated from the original CBD model. The fan charts derived from the LLCBD model encompass a larger collection of possible long-term mortality scenarios, ranging from a zero mortality improvement rate to improvement rates that are even greater that those realized in the recent past. As explained below, we view the provision of fan charts that are wider in the long run as an advantage.

When developing a mortality model, it is important to consider biological reasonableness, a concept that was first raised by Cairns et al. (2006) in the context of median mortality forecasts. Simply put, this concept means that the collective views of experts in mortality should be taken into account. For example, one should rule out a model that projects a strictly positive probability of immortality.

The concept of biological reasonableness should also be applicable to interval forecasts. In the context interval mortality forecasting, we believe that it is legitimate to interpret biological reasonableness as follows: if the collective view of experts is that future life expectancies should not exceed a certain range, then a biologically reasonable interval forecast should not be wider than the range possible of outcomes that experts agree on; in contrast, if experts have rather different opinions on the prospect of longevity, then a biologically reasonable interval forecast should take a shape that encompasses as much as possible the range of opinions. We use two examples to explain why we regard the interval mortality forecasts produced by our LLCBD model are more biologically reasonable.

The seminal work of Oeppen and Vaupel (2002) found that the trend in record life expectancy since 1840 is close to perfectly linear, showing no sign of deceleration. It was then argued that the linear trend would continue in the coming decades. To achieve a linear climb in life expectancy, age-specific death probabilities need to decline at an increasing pace. As shown in Figure 3.7, the plausibility of an increasing rate of mortality recline can be captured in the fan charts derived from the LLCBD model, but not in those generated from the original CBD model.

The recent reports produced by the Society of Actuaries (2014) and the Canadian Institute of Actuaries (2014) suggest that (at least part of) the actuarial profession in North America believes that mortality improvement rates will reduce to 0-1% after a transitional period of some 20-30 years. The profession's view means that trajectories of future mortality rates will become flat or almost flat a few decades from now. Such an outcome does not seem to be captured by the fan charts generated using the original CBD model, but may possibly be contained in the LLCBD fan charts whose widths increase with time more quickly.

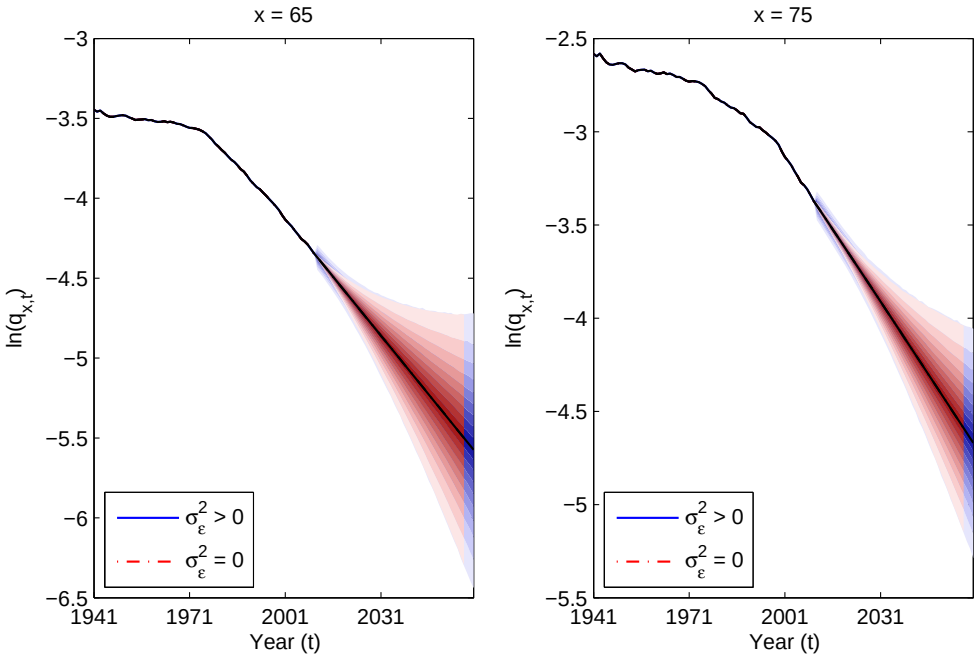


Figure 3.8: Forecasts of $\ln(q_{x,t})$ at $x = 65, 75$ produced by the LLCBD models with $\sigma_\epsilon^2 = 0$ (excluding variation in death counts) and with $\sigma_\epsilon^2 > 0$ (including variation in death counts).

3.3.6 Excluding Variation in Death Counts

As the quantity being modeled ($q_{x,t}$) by the LLCBD model is the *crude* conditional death probability, the fan charts in Figure 3.7 incorporate both systematic longevity risk and variation in the actual number of deaths. They reflect the level of uncertainty surrounding the future *crude* death probabilities, assuming that the population size remains stable in the future.

To exclude the uncertainty due to variation in the actual number of deaths, we can generate fan charts of future death probabilities under the assumption that $\sigma_\epsilon^2 = 0$. Figure 3.8 compares the fan charts that incorporate only systematic longevity risk with those that incorporate both sources of uncertainty. It can be observed that systematic longevity risk accounts for most of the total uncertainty. This outcome is not overly surprising, because the population being modeled is a national population with a reasonably large number of persons-at-risk at each age over the age range under consideration.

3.3.7 Further Comments on the Dynamics of $C_1(t)$ and $C_2(t)$

We assume that $C_i(t)$, $i = 1, 2$ follows a random walk rather than a mean-reverting stationary process (e.g., an AR(1)), on grounds that we have no a priori knowledge about the value and more importantly the existence of the mean of the drift vector. To substantiate the random walk assumption with statistical evidence, we consider the Dickey-Fuller test, which tests the null hypothesis of

$$z(t) = z(t-1) + \epsilon_z(t)$$

against the alternative hypothesis of

$$z(t) = \mu_z + \phi_z z(t-1) + \epsilon_z(t)$$

for a generic time-series $\{z(t)\}$, where μ_z is a constant and ϕ_z is another constant with an absolute value that is strictly smaller than 1. The test is applied to the following:

(a) 5-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ in the original CBD model, estimated using the least squares method;

(b) 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ in the original CBD model, estimated using the least squares method;

(c) the retrieved values of $C_1(t)$ and $C_2(t)$ in the LLCBD model.

As mentioned in Section 3.2, (a) and (b) may be regarded as proxies for the drifts at different time points. The results of all tests performed (see Table 3.5) are in favour of a random walk.

3.4 Other Modeling Considerations

3.4.1 A Comparison with Models with Additional Dynamic Factors and/or Age Effect Structures

One may wonder if the benefits of using a stochastic drift process can be achieved by using models with additional dynamic factors and/or different age effect structures. For this reason, we further compare the proposed LLCBD model against the following four

Table 3.5: The results of the Dickey-Fuller tests for a random walk against an AR(1), applied to the 5- and 10-year moving averages of $\Delta\kappa_1(t)$ and $\Delta\kappa_2(t)$ in the original CBD model and the retrieved values of $C_1(t)$ and $C_2(t)$ in the LLCBD model. A 5% level of significance is used.

p-value	Test statistic	Critical value	p-value	Test statistic	Critical value
5-year moving averages of $\Delta\kappa_1(t)$			5-year moving averages of $\Delta\kappa_2(t)$		
0.6293	-1.2349	-2.9097	0.1596	-2.3529	-2.9097
10-year moving averages of $\Delta\kappa_1(t)$			10-year moving averages of $\Delta\kappa_2(t)$		
0.9572	0.0316	-2.9141	0.5815	-1.3441	-2.9141
Retrieved values of $C_1(t)$			Retrieved values of $C_2(t)$		
0.9802	0.3685	-2.9054	0.7115	-1.0467	-2.9054

discrete-time stochastic mortality models, which have been considered extensively in the literature (see, e.g., Cairns et al., 2009, 2011a; Dowd et al., 2010a,b).[6]

- The original Lee-Carter model (Model M1):

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

  where a_x and b_x are age-specific parameters, k_t is a time-varying dynamic factor, and $\epsilon_{x,t}$ is the sampling error. Compared to the CBD/LLCBD model, Model M1 has one fewer dynamic factor and a different age effect structure (specified by parameters a_x and b_x).

- The Renshaw-Haberman model (Model M2):

$$\ln(m_{x,t}) = a(x) + b_1(x)\kappa(t) + b_2(x)\gamma(t-x) + \epsilon_{x,t},$$

  where $a(x)$, $b_1(x)$ and $b_2(x)$ are age-specific parameters, $\kappa(t)$ is a time-varying dynamic factor, $\gamma(t-x)$ is a cohort-varying dynamic factor, and $\epsilon_{x,t}$ is the sampling error. Compared to the CBD/LLCBD model, Model M2 has the same number of dynamic factors, but one of which is cohort-related rather than time-related. It has also a different age effect structure.

- The Cairns-Blake-Dowd model with a cohort effect (Model M6):

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_1(t) + \kappa_2(t)(x-\bar{x}) + \gamma(t-x) + \epsilon_{x,t},$$

  where $\bar{x}$ is the average of the age range $[x_a, x_b]$ to which the model is calibrated, $k_1(t)$ and $k_2(t)$ are time-varying dynamic factors, $\gamma(t-x)$ is a cohort-varying dynamic factor, and $\epsilon_{x,t}$ is the sampling error. Compared to the CBD/LLCBD model, Model M6 contains one extra dynamic factor, which varies with year-of-birth.

- The Cairns-Blake-Dowd model with quadratic and cohort effects (Model M7):

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(4)} + \epsilon_{x,t},$$

[6] We do not consider Model M3 (the age-period-cohort model), because it is simply a special case of Model M2 with $b_1(x)$ and $b_2(x)$ being constants instead of functions of age. Also, Model M4 (the P-splines regression) is excluded, in part because it is based on a regression rather than stochastic processes and in part because it does not yield sample mortality paths which are needed for the analyses in later parts of the chapter.

where $\hat{\sigma}_x^2$ is the mean of $(x-\bar{x})^2$ over $[x_a, x_b]$. Compared to the CBD/LLCBD model, Model M7 contains two extra dynamic factors, one of which varies with time and the other of which varies with year-of-birth.

As what we did for the CBD and LLCBD models in Section 3.3.4, we evaluate the forecasting performance of the four additional models by

(i) using the 'contracting horizon backtest' considered in Dowd et al. (2010b), and

(ii) fitting the models to restricted calibration windows and then comparing the resulting forecasts with the actual values that are not used in fitting the models.

The result of (i) is shown in Figure 3.9. By comparing Figure 3.9 with the upper panel of Figure 3.4, we can conclude that none of the four additional models can produce forecasts with the desirable properties (more accurate median forecasts and more adequate provisions of uncertainty) possessed by the LLCBD forecasts.

The result of (ii) is reported in Figure 3.10. The LLCBD model generally yields ME and MSE with smaller magnitudes compared to any one of the alternative models. The benefit of using the LLCBD model is the most apparent when the end point of the calibration window is between 1980 and 1990.

The additional evaluation work indicates that the benefit of using a stochastic drift process cannot be obtained simply by using more dynamic factors or tweaking the age-effect structures. Of course, it may be possible to further improve the forecasting performance by adding more dynamic factors (e.g., a cohort effect) to the LLCBD model. These possible extensions are left for future research.

3.4.2 Sensitivity to the Choice of Age Range

The baseline estimation result is based on an age range of $[x_a, x_b] = [50, 89]$. This age range is chosen for the following reasons.

First, the age range of [50, 89] is often used in the literature (see, e.g., Cairns et al., 2009) to calibrate stochastic mortality models for pension and annuity valuations. Using this age range enables readers to compare the estimation results in this chapter and other papers more readily.

Second, according to the Human Mortality Database documentation (Wilmoth et al., 2005), raw population counts for Canadians are available up to age 89 only. Population counts beyond age 89 are not 'real' but estimated using the extinct cohort method.

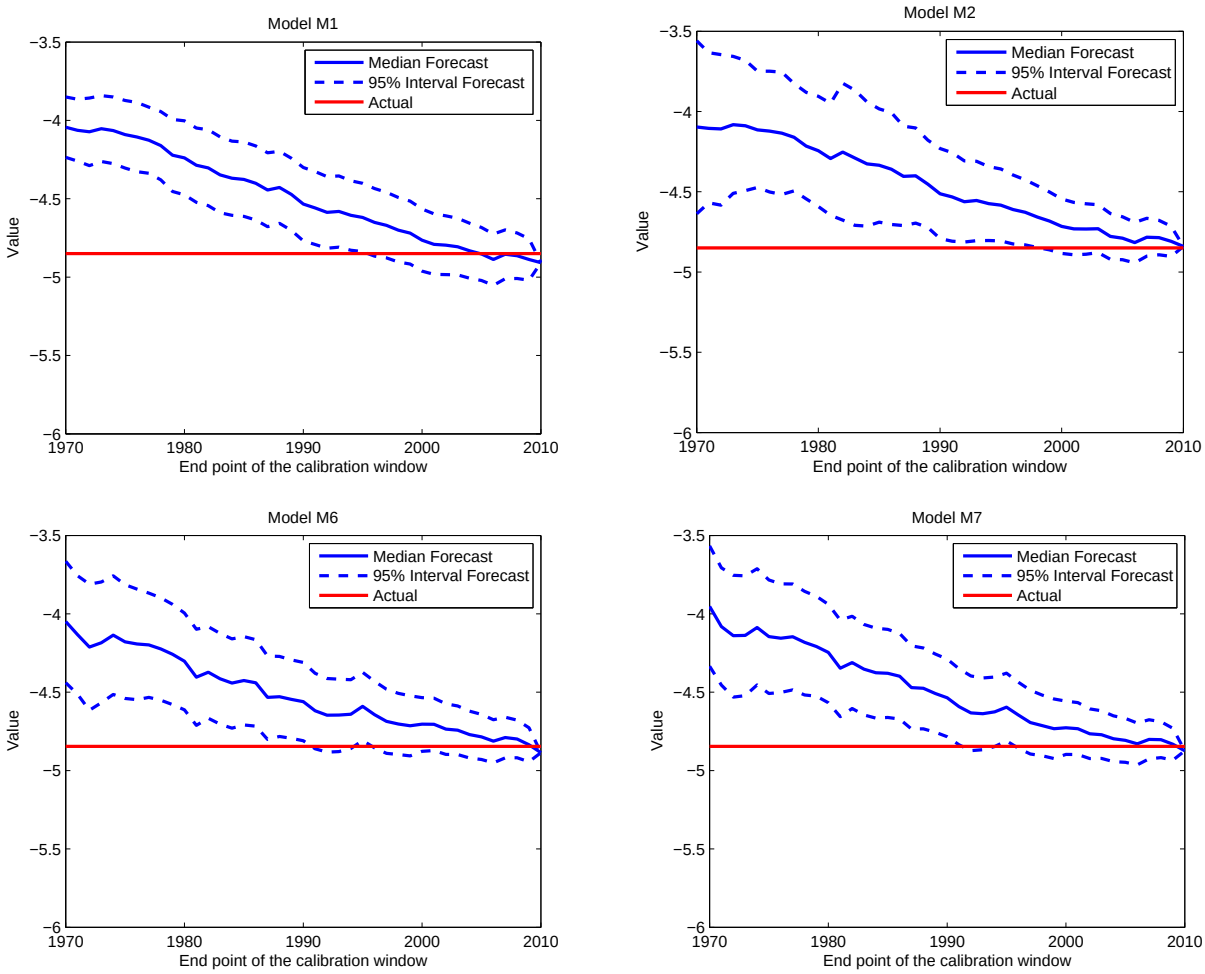


Figure 3.9: The median and 95% interval forecasts of $\ln(q_{60,2010}/1 - q_{60,2010})$, generated from Models M1, M2, M6 and M7 that estimated to data over different calibration windows. The starting point of the calibration windows is always 1941 but end points range from 1971 to 2009.

Third, as the models are built for modeling longevity risk at pensionable ages, beginning the sample age range at 50 prevents the models from being influenced by the (possibly different) mortality improvement dynamics at younger ages. Also, the age effect structure in the CBD/LLCBD model does not capture the accident hump at younger ages.

In this sub-section, we examine how the estimation results may be different if different age ranges are used. We consider the following six age ranges: 45-94, 46-93, 47-92, 48-91, 49-90 and 50-89 (the baseline age range). Note that the average age over each of the six

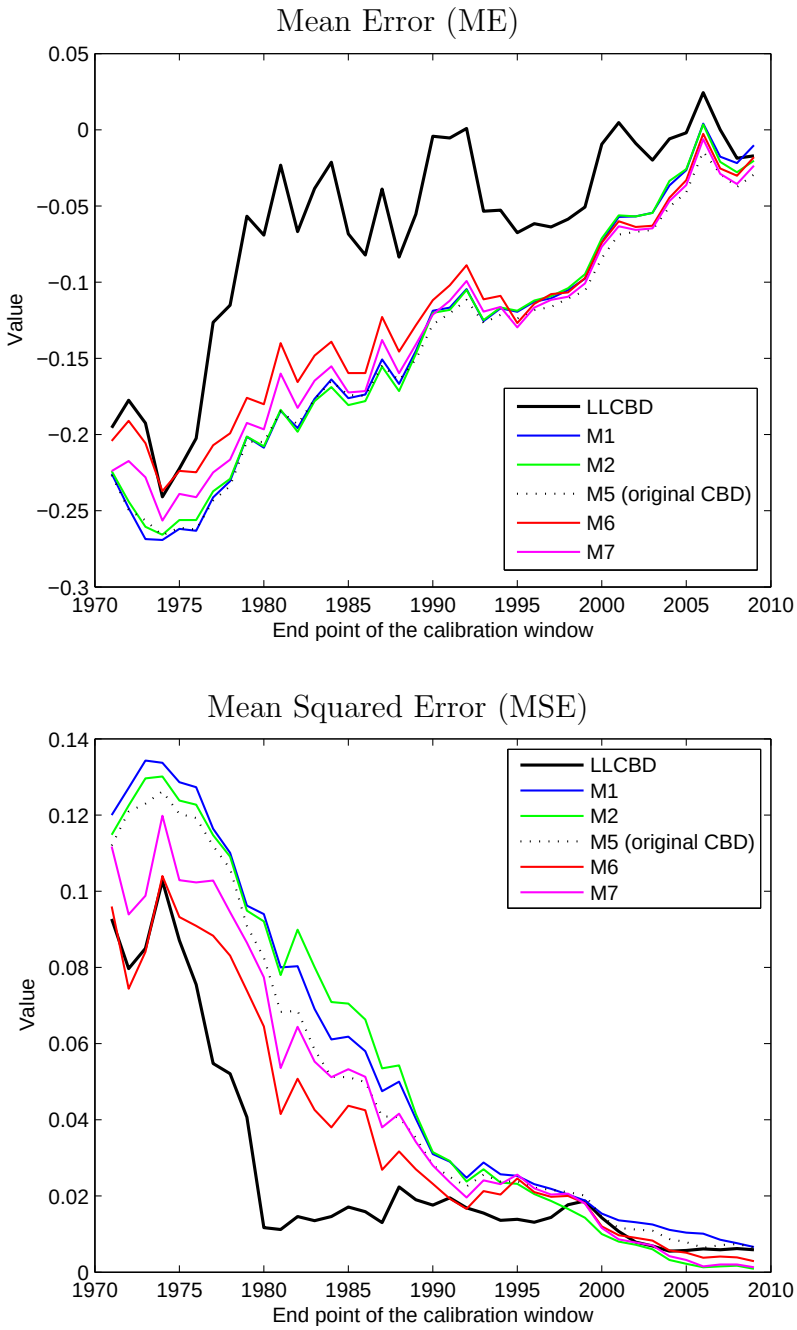


Figure 3.10: The Mean Error (ME) and Mean Squared Error (MSE) for the forecasts of $\ln(q_{x,t}/1 - q_{x,t})$ produced by the LLCBD model and Models M1, M2, M5 (the original CBD), M6 and M7.

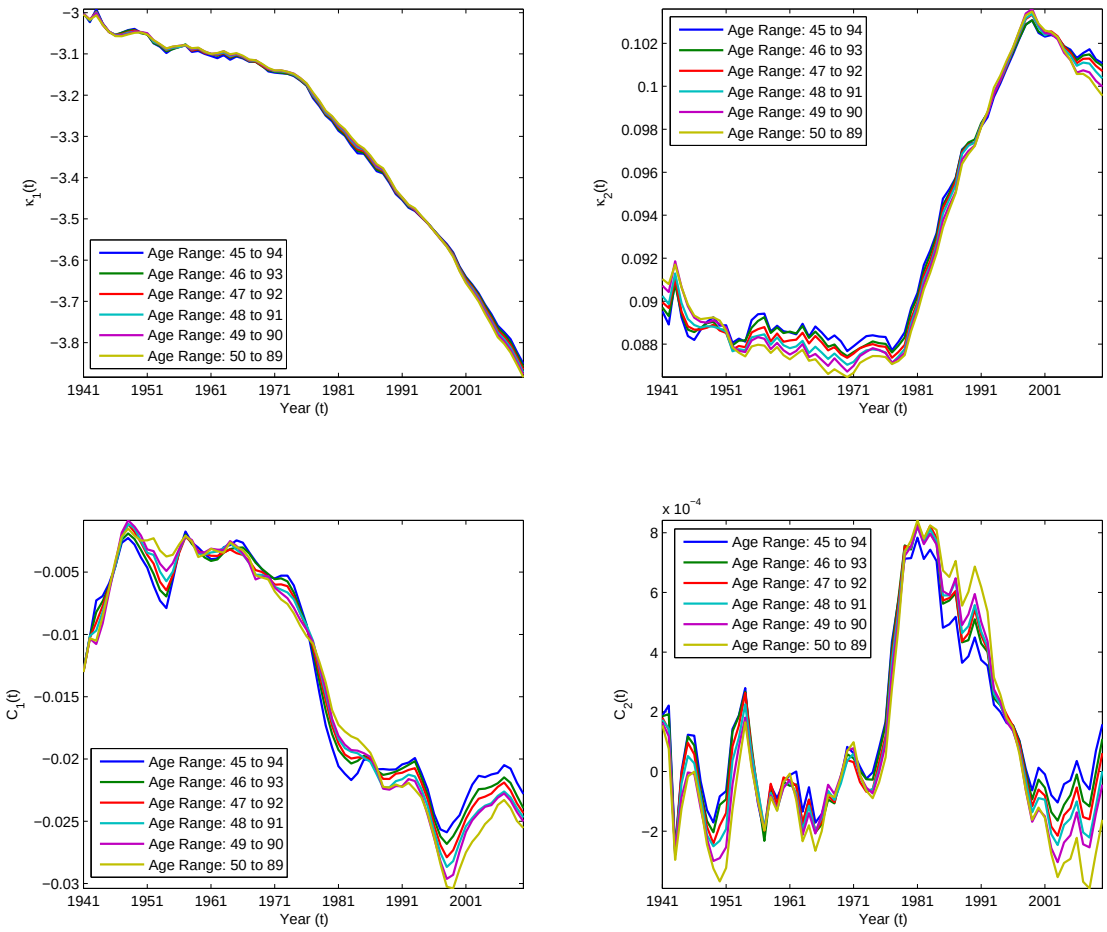


Figure 3.11: The retrieved values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model when different age ranges are used in estimation.

age ranges is $\bar{x} = 69.5$. It is necessary to keep $\bar{x}$ fixed, because otherwise $\kappa_1(t)$ and $\kappa_2(t)$ would be scaled differently.

Figure 3.11 shows the retrieved values of the hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, in the LLCBD model for each age range under consideration. For most of the time, the retrieved hidden states are robust relative to the choice of age range. Over the last 10 years of the calibration window, the retrieved drifts are arguably quite sensitive to the age range used. However, compared to the widths of the confidence intervals (Figure 3.3), the changes in the retrieved drifts due to changes in age range are small.

Table 3.6: The LMPI test results (test statistic and critical values at 5% and 10% levels of significance) for $\kappa_1(t)$ and $\kappa_2(t)$ estimated from the four additional data sets.

Data set	Test statistic		Critical value	
	$\kappa_1(t)$	$\kappa_2(t)$	5%	10%
Dutch male	1.4714	0.5775	0.4709	0.3503
English and Welsh male	0.5909	0.0312	0.4689	0.3485
Japanese unisex	0.0450	0.5258	0.4709	0.3503
Canadian female	0.0210	0.0567	0.4689	0.3485

3.4.3 Application to other Data Sets

It would be interesting to see if stochastic drifts apply to the mortality dynamics of other populations, and if the LLCBD model still outperforms when it is fitted to other data sets. In this sub-section, we apply the testing and modeling methods to the following additional data sets:

Population	Age range	Sample period
Dutch male	50 to 89	1951 to 2010
English and Welsh male	50 to 89	1941 to 2010
Japanese unisex	50 to 89	1951 to 2010
Canadian female	50 to 89	1941 to 2010

These data sets cover different geographical locations (Europe, North America and Asia), genders (male, female and unisex) and sample periods (1941-2010 and 1951-2010). They are also obtained from the Human Mortality Database.

First, we apply the LMPI test to the least square estimates of the CBD dynamic factors $\kappa_1(t)$ and $\kappa_2(t)$. The test results are reported in Table 3.6. For Dutch males, the test results indicate that the drifts for both $\kappa_1(t)$ and $\kappa_2(t)$ are stochastic. More interestingly, the test results suggest that it is possible to have only one drift being stochastic ($\kappa_1(t)$ for English and Welsh males and $\kappa_2(t)$ for Japanese unisex). For Canadian females, the test results conclude that none of the two drifts is stochastic. This conclusion highlights a noteworthy fact: just as cohort effect (which is highly significant in the United Kingdom but not so much in Asian populations) and jump effect (which can only be detected if the calibration window is long enough), stochastic drift is not a universal phenomenon. Whether a stochastic drift is needed depends critically on the data set considered.

Next, we use the AIC to compare the goodness-of-fit produced by the LLCBD models (with one and two stochastic drift(s)) and the original CBD model. The results are shown

Table 3.7: The values of $\mathcal{N}$, $\ln(\hat{\mathcal{L}})$ and AIC for the models fitted to the four additional data sets. ('CBD': the original CBD model; 'LLCBD': the LLCBD model with two stochastic drifts; 'LLCBD*': the LLCBD model with C_2 being constant; 'LLCBD**': the LLCBD model with C_1 being constant.)

	Dutch Male			English and Welsh Male		
Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	3547.7047	−7087.4093	4	4146.5332	−8285.0664	4
LLCBD*	3556.3097	−7098.6194	7	4149.9301	**−8285.8603**	7
LLCBD**	3554.6960	−7095.3920	7	4149.4382	−8284.8764	7
LLCBD	3562.7838	**−7103.5676**	11	4153.3105	−8284.6209	11
	Japanese Unisex			Canadian Female		
Model	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$	$\ln(\hat{\mathcal{L}})$	AIC	$\mathcal{N}$
CBD	2662.9849	−5317.9698	4	3304.2798	**−6600.55951**	4
LLCBD*	2666.4540	−5318.9079	7	3305.5235	-6597.04696	7
LLCBD**	2671.2966	**−5328.5931**	7	3306.0631	−6598.12613	7
LLCBD	2671.7738	−5321.5476	11	3307.2052	−6592.41038	11

in Table 3.7. For Dutch males, the optimal model is the LLCBD model with two stochastic drifts; for English and Welsh males, the optimal model is the LLCBD model with C_2 being constant; for Japanese unisex, the optimal model is the LLCBD model with C_1 being constant; and for Canadian females, the optimal model is the original CBD model. These conclusions are in line with the LMPI test results.

For all of the four populations except Canadian females, we compare the forecasts generated by the original CBD model and the model chosen according to the AIC. Figure 3.12 displays the results of the 'contracting horizon backtest' (used in Sections 3.3.4 and 3.4.1) for the three populations. In all cases, the LLCBD model (or its variant) yields a median forecast that is less biased and a 95% prediction interval that captures a larger proportion of the actual values.

Figure 3.13 shows the values of ME and MSE produced by the 'forecasts' generated from models that are estimated to restricted calibration windows (1941-1971, ..., 1941-2009 for English and Welsh males; 1951-1971, ..., 1951-2009 for Dutch males and Japanese unisex). For Dutch males, the LLCBD model – which contains two stochastic drifts – yields significantly higher forecast accuracy compared to the original CBD model. For the other two populations, the LLCBD models – of which the drifts are only partially stochastic – still provide improved forecast accuracy, but the improvement is not that remarkable.

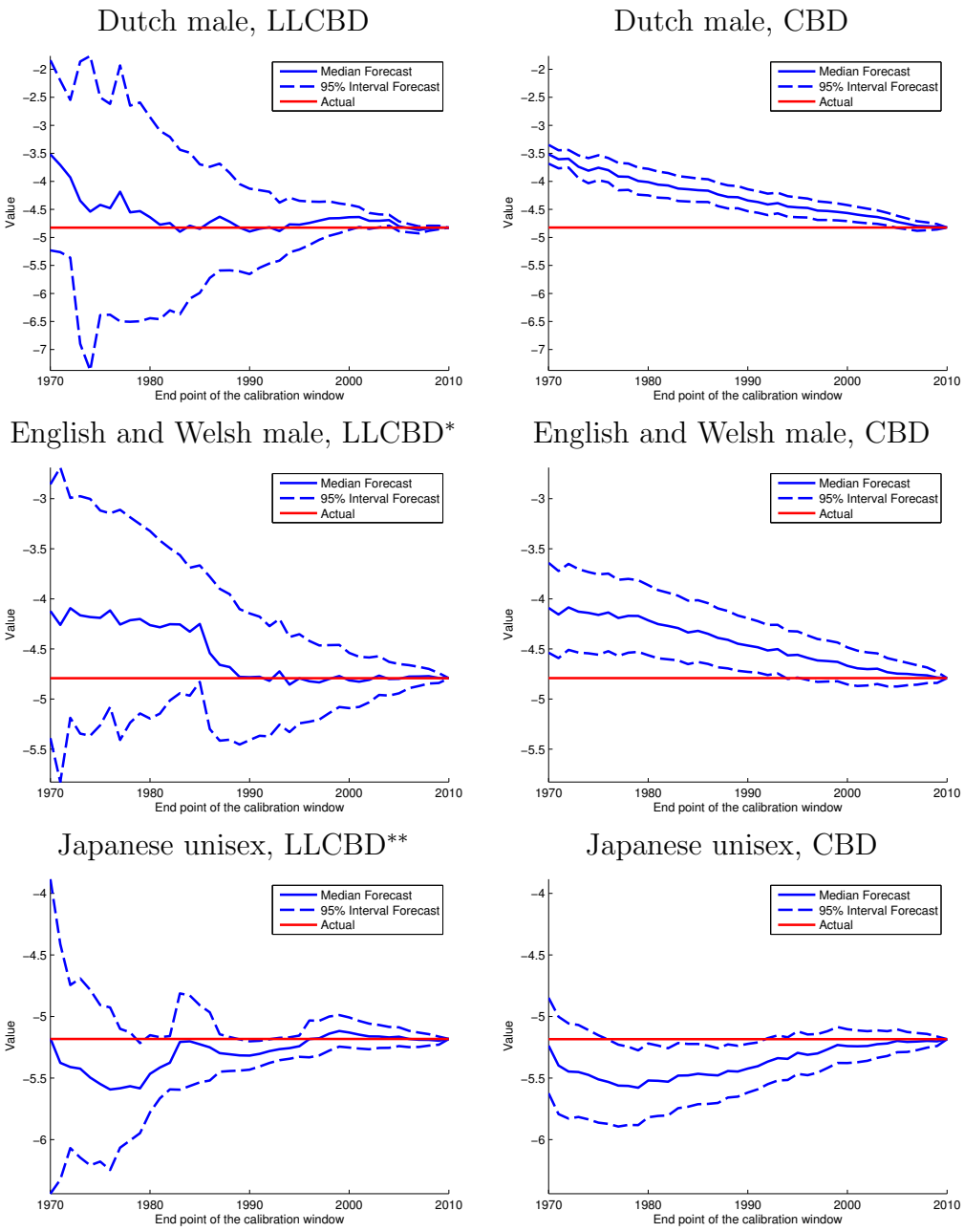


Figure 3.12: The median and 95% interval forecasts of $\ln(q_{60,2010}/1 - q_{60,2010})$, generated from models that estimated to data over different calibration windows. The starting point of the calibration windows is either 1951 (Dutch male and Japanese unisex) or 1941 (English and Welsh male), but end points range from 1971 to 2009.

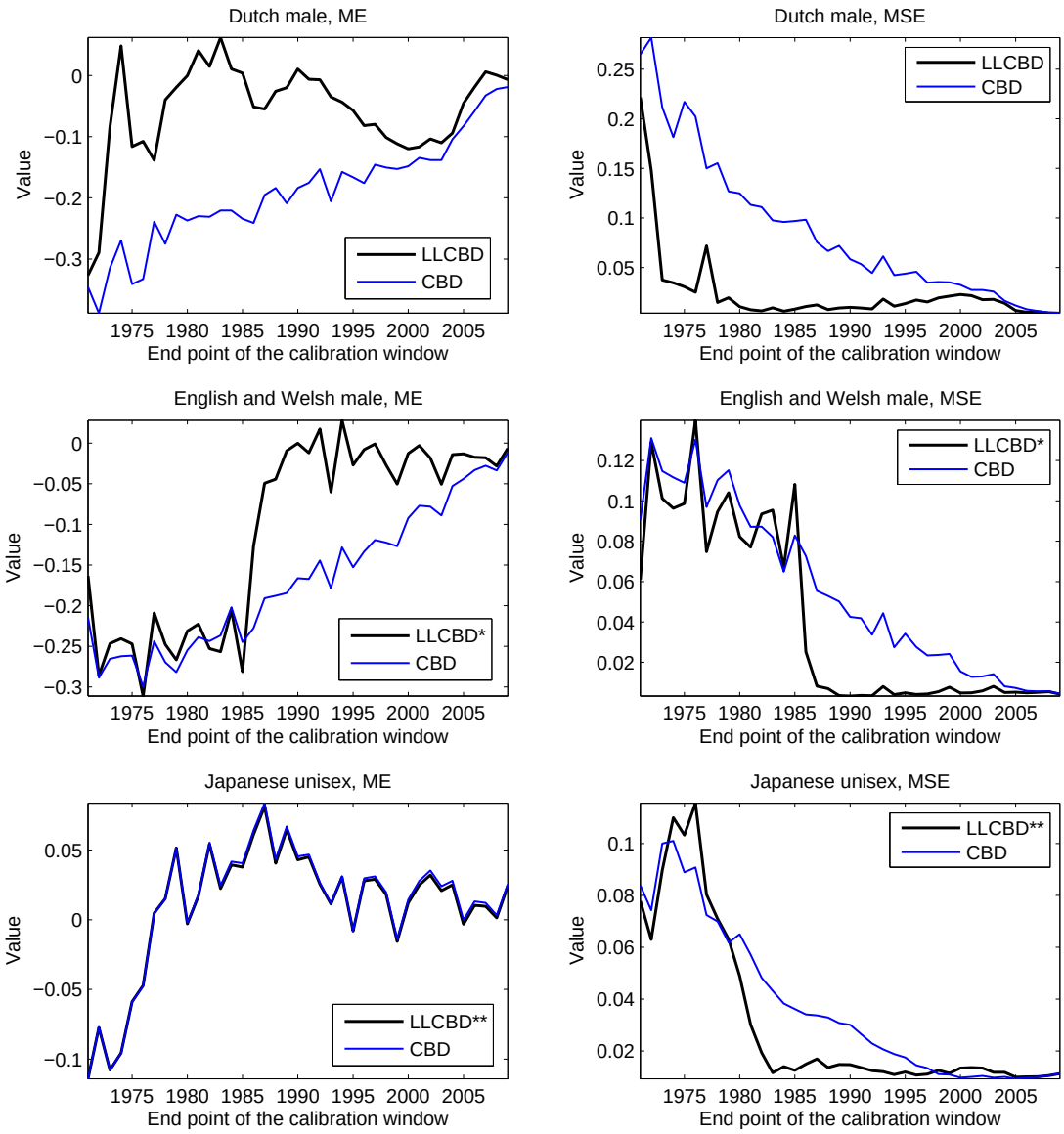


Figure 3.13: The Mean Error (ME) and Mean Squared Error (MSE) for the forecasts of $\ln(q_{x,t}/(1-q_{x,t})$ produced by the original CBD model and the optimal LLCBD model.

Finally, we evaluate the robustness of the models relative to the length of the calibration window used. The results are provided in Figure 3.14. As what we found in Sections 3.3.4 and 3.4.1, the LLCBD model (or its variant) is more robust than the original CBD model. It also yields forecasts that are more consistent with the recent trend.

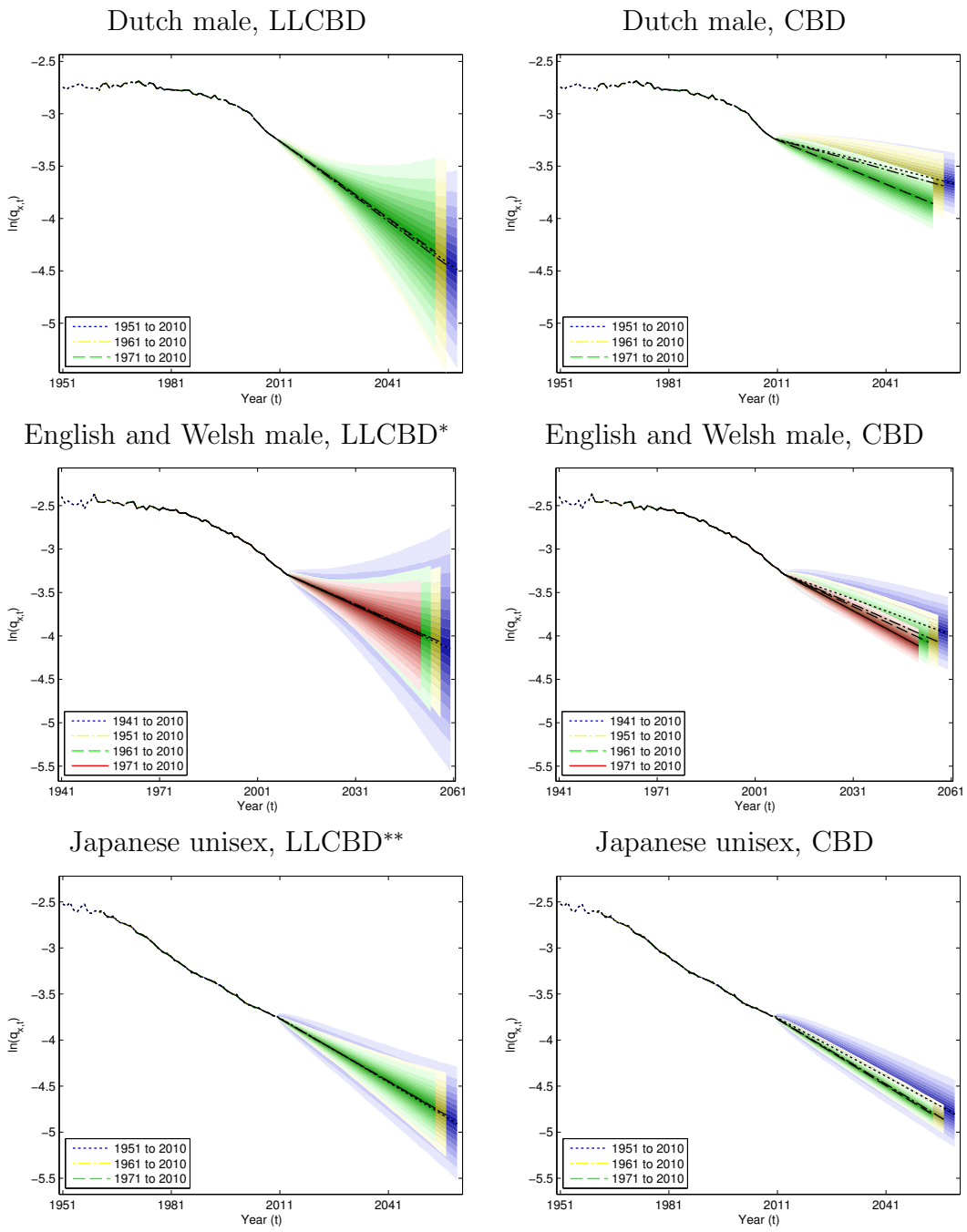


Figure 3.14: Forecasts of $\ln(q_{x,t})$ for $x = 75$ produced by the original CBD model and the optimal LLCBD model fitted to data over different calibration windows.

3.5 Hedging Drift and Diffusion Risks

Having modeled both drift and diffusion risks with the LLCBD model, in this section we explain how we may hedge these risks by using standardized hedging instruments. We start with a sub-section which details the assumptions made, followed by a review of the traditional delta and delta-nuga hedging methods. We then introduce our proposed 'generalized state-space hedging method', and conclude with some comments about the distinctions between the traditional and newly proposed methods.

3.5.1 The Set-up

Our goal is to hedge the longevity risk associated with a T-year temporary life annuity immediate that is just sold. Suppose that it is now time t_0 (i.e., the end of year t_0) and that the annuitant is now aged x_0. Ignoring sampling risk, the (random) present value of the liability being hedged is

$$L = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \tilde{p}_{x_0+t-t_0-1,t}$$

per contract, where r is the interest rate for discounting purposes,

$$\tilde{p}_{x,t} = 1 - \tilde{q}_{x,t}$$

and, as defined in equation (3.3), $\tilde{q}_{x,t}$ is the underlying (unobserved) conditional probability of death in year t (between age x and $x+1$). Note that L is a function of $\kappa_1(t_0+1), \ldots, \kappa_1(t_0+T)$ and $\kappa_2(t_0+1), \ldots, \kappa_2(t_0+T)$, all of which are random as of the time when the longevity hedge is established.

The standardized hedging instruments used are q-forwards. A q-forward is a zero-coupon swap with its floating leg proportional to the realized death probability at a certain reference age in a certain reference year and its fixed leg proportional to the corresponding pre-determined forward mortality rate.

We consider a hedge portfolio of $m \geq 1$ q-forwards. Let x_j, t_j and $q^f_{x_j,t_j}$ be the reference age, reference year and forward mortality rate for the jth q-forward contract, respectively. We assume $t_0 < t_1 < \ldots < t_m$. We also assume that payment exchanges hands at the end of the reference year, so that $t_j - t_0$ is the time-to-maturity for the jth q-forward. For simplicity, it is assumed in the derivations that there is no population basis risk; that is, the q-forwards and the liability being hedged are associated with exactly the same population

of individuals. In today's market, the LLMA's[7] LifeMetrics indexes, to which standardized q-forwards are linked, are based on smoothed (rather than crude) age-specific conditional death probabilities. For this reason, we assume that the q-forwards are linked to the underlying death probabilities that are not subject to the randomness of $\epsilon_{x,t}$. Under the mentioned assumptions, the (random) present value of the payoff from the jth q-forward contract is given by

$$H_j = e^{-r(t_j - t_0)}(q^f_{x_j,t_j} - \tilde{q}_{x_j,t_j}), \quad j = 1, \ldots, m.$$

It is obvious that H_j depends on $\kappa_1(t_j)$ and $\kappa_2(t_j)$, which are random variables as of the time when the longevity hedge is established.

We let N_j be the notional amount of the jth q-forward and

$$P = L - \sum_{j=1}^{m} N_j H_j$$

be the (random) present value of all cash flows when the longevity hedge is in place. If the longevity hedge is successful, then the variability in P would be significantly less than that in L. Using this reasoning, we assess hedge effectiveness with the following metric:

$$HE = 1 - \frac{\text{Var}(P)}{\text{Var}(L)}.$$

The value of HE is close to 1 if the longevity hedge is effective, and close to 0 if otherwise. This metric is also used by Cairns (2011, 2013), Cairns et al. (2014), Coughlan et al. (2011) and Li and Hardy (2011).

3.5.2 A Review of Traditional Delta and Delta-Nuga Hedging Methods

3.5.2.1 Delta Hedging

The idea behind the traditional delta hedging method, considered by researchers including Cairns (2011, 2013) and Zhou and Li (2014), is to match the sensitivities of the liability being hedged and the hedging portfolio with respect to changes in the time-varying factors in the assumed stochastic mortality model in year t_0. If the CBD model is assumed, then

[7]The Life and Longevity Markets Association (www.llma.org)

the sensitivities involved are represented by the following partial derivatives: $\partial L/\partial\kappa_i(t_0)$, $\partial H_j/\partial\kappa_i(t_0)$, $i = 1, 2$, $j = 1, \ldots, m$.

Because L and H_j are functions of $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$ rather than $\kappa_1(t_0)$ and $\kappa_2(t_0)$, the partial derivatives of L and H_j with respect to $\kappa_1(t_0)$ and $\kappa_2(t_0)$ cannot be computed straightforwardly. To make the estimation of sensitivities possible, the partial derivatives are calculated using the best estimates of L and H_j on the basis of $\kappa_1(t_0)$ and $\kappa_2(t_0)$:

$$\hat{L} = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \quad \text{and} \quad \hat{H}_j = e^{-r(t_j-t_0)}(q^f_{x_j,t_j} - \hat{q}_{x_j,t_j}), \tag{3.5}$$

where $\hat{q}_{x,t} = \frac{e^{\hat{\kappa}_1(t)+(x-\bar{x})\hat{\kappa}_2(t)}}{1+e^{\hat{\kappa}_1(t)+(x-\bar{x})\hat{\kappa}_2(t)}}$, $\hat{p}_{x,t} = 1 - \hat{q}_{x,t}$ and

$$\hat{\kappa}_i(t) = \kappa_i(t_0) + C_i(t_0) \times (t - t_0), \quad t > t_0, \quad i = 1, 2,$$

is the best estimate of $\kappa_i(t)$ given $\kappa_i(t_0)$ and $C_i(t_0)$.[8]

The partial derivatives of $\hat{L}$ and $\hat{H}_j$ with respect to $\kappa_1(t_0)$ and $\kappa_2(t_0)$ can be computed readily as follows:

$$\begin{aligned}
\frac{\partial\hat{L}}{\partial\kappa_1(t_0)} &= -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} \hat{q}_{x_0+t-t_0-1,t}\right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t}\right); \\
\frac{\partial\hat{L}}{\partial\kappa_2(t_0)} &= -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (x_0+t-t_0-1-\bar{x})\hat{q}_{x_0+t-t_0-1,t}\right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t}\right); \\
\frac{\partial\hat{H}_j}{\partial\kappa_1(t_0)} &= -e^{-r(t_j-t_0)}\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}; \\
\frac{\partial\hat{H}_j}{\partial\kappa_2(t_0)} &= -e^{-r(t_j-t_0)}(x_j-\bar{x})\hat{p}_{x_j,t_j}\hat{q}_{x_j,t_j}.
\end{aligned}$$

These derivatives are regarded as the 'deltas' of the liability being hedged and the hedging instruments.

We require exactly two hedging instruments to obtain a delta-neutral position. The notional amounts of the two q-forwards in the hedge portfolio should satisfy the following system of equations:

$$\begin{cases} \frac{\partial\hat{L}}{\partial\kappa_1(t_0)} = N_1 \times \frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} + N_2 \times \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)} \\ \frac{\partial\hat{L}}{\partial\kappa_2(t_0)} = N_1 \times \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)} + N_2 \times \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)} \end{cases},$$

[8] In effect, $\hat{L}$ and $\hat{H}_j$ are respectively the values of L and H_j calculated by switching off all random components.

the solution to which is given by

$$\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{L}}{\partial \kappa_2(t_0)} \end{pmatrix},$$

provided that the inverse of the square matrix on the right-hand-side exists. The invertibility of the square matrix is discussed in Section 3.5.5.

3.5.2.2 Delta-Nuga Hedging

Cairns (2013) acknowledged that the values of the liability being hedged and the hedge portfolio are also affected by changes in the estimated values of the drift terms in the processes for $\kappa_1(t)$ and $\kappa_2(t)$. To mitigate this piece of uncertainty, he proposed the delta-nuga hedging method in which the sensitivities of $\hat{L}$ and $\hat{H}_j$ to the changes in the drift terms are also considered. This method includes, additionally, the following four partial derivatives:

$$\begin{aligned}
\frac{\partial \hat{L}}{\partial C_1(t_0)} &= -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (t-t_0) \hat{q}_{x_0+t-t_0-1,t} \right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right); \\
\frac{\partial \hat{L}}{\partial C_2(t_0)} &= -\sum_{u=1}^{T} e^{-ru} \left(\sum_{t=t_0+1}^{t_0+u} (t-t_0)(x_0+t-t_0-1-\bar{x}) \hat{q}_{x_0+t-t_0-1,t} \right) \left(\prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0+t-t_0-1,t} \right); \\
\frac{\partial \hat{H}_j}{\partial C_1(t_0)} &= -e^{-r(t_j-t_0)} (t_j - t_0) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}; \\
\frac{\partial \hat{H}_j}{\partial C_2(t_0)} &= -e^{-r(t_j-t_0)} (t_j - t_0)(x_j - \bar{x}) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}.
\end{aligned}$$

These additional partial derivatives are considered as the 'nugas' of the liability being hedged and the hedging instruments. To neutralize the deltas and nugas, we need exactly four hedging instruments. The notional amounts of the four hedging instruments should satisfy the following system of equations:

$$\begin{cases}
\frac{\partial \hat{L}}{\partial \kappa_1(t_0)} = N_1 \times \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} + N_2 \times \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} + N_3 \times \frac{\partial \hat{H}_3}{\partial \kappa_1(t_0)} + N_4 \times \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\
\frac{\partial \hat{L}}{\partial \kappa_2(t_0)} = N_1 \times \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} + N_2 \times \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} + N_3 \times \frac{\partial \hat{H}_3}{\partial \kappa_2(t_0)} + N_4 \times \frac{\partial \hat{H}_4}{\partial \kappa_2(t_0)} \\
\frac{\partial \hat{L}}{\partial C_1(t_0)} = N_1 \times \frac{\partial \hat{H}_1}{\partial C_1(t_0)} + N_2 \times \frac{\partial \hat{H}_2}{\partial C_1(t_0)} + N_3 \times \frac{\partial \hat{H}_3}{\partial C_1(t_0)} + N_4 \times \frac{\partial \hat{H}_4}{\partial C_1(t_0)} \\
\frac{\partial \hat{L}}{\partial C_2(t_0)} = N_1 \times \frac{\partial \hat{H}_1}{\partial C_2(t_0)} + N_2 \times \frac{\partial \hat{H}_2}{\partial C_2(t_0)} + N_3 \times \frac{\partial \hat{H}_3}{\partial C_2(t_0)} + N_4 \times \frac{\partial \hat{H}_4}{\partial C_2(t_0)}
\end{cases},$$

which implies

$$\begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \frac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix}, \tag{3.6}$$

provided that the inverse of the square matrix on the right-hand-side exists. The invertibility of the square matrix is discussed in Section 3.5.5.

3.5.3 The Generalized State-Space Hedging Method

The review presented in the previous sub-section exposes two limitations of the traditional delta and delta-nuga hedging methods. First, in deriving the deltas and nugas, it is assumed that $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$ are linear functions of $\kappa_1(t_0)$, $\kappa_2(t_0)$, $C_1(t_0)$ and $C_2(t_0)$. The resulting deltas and nugas therefore contain no information about the sensitivities of L and H_j to $\kappa_1(t)$ and $\kappa_2(t)$ for any $t > t_0$. Second, the traditional methods have very stringent requirements on the number of hedging instruments being used. The delta hedging method requires exactly two distinct hedging instruments whereas the delta-nuga hedging needs exactly four.

To overcome these limitations, we hereby introduce the generalized state-space hedging method which better utilizes the information contained in the hidden states at different times and is more flexible in terms of the number of hedging instruments required. In the generalized state-space hedging method, we work on L and H_j (rather than their best estimates) and preserve the fact that they are functions of $\kappa_1(t)$ and $\kappa_2(t)$ for $t > t_0$.

We let $\vec{\alpha}_t^* = (\kappa_1(t), \kappa_2(t))'$ and $\vec{\alpha}_t^{**} = (C_1(t), C_2(t))'$. It follows that L is a function of the sequence of $\{\vec{\alpha}_{t_0+1}^*, \ldots, \vec{\alpha}_{t_0+T}^*\}$ and H_j is a function of $\vec{\alpha}_{t_j}^*$, $j = 1, \ldots, m$. The derivation of the generalized state-space hedging strategy involves the first-order Taylor approximations of L and H_j about all relevant vectors of hidden states. For L, the first-order approximation $l(\vec{\alpha}_{t_0+1}^*, \ldots, \vec{\alpha}_{t_0+T}^*)$ is given by

$$L \approx l(\vec{\alpha}_{t_0+1}^*, \ldots, \vec{\alpha}_{t_0+T}^*) = \hat{L} + \sum_{i=t_0+1}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*} \right)' (\vec{\alpha}_i^* - \hat{\vec{\alpha}}_i^*),$$

where $\hat{L}$ is defined in equation (3.5) and $\hat{\vec{\alpha}}_i^*$ is the expected value of $\vec{\alpha}_i^*$ given the information up to and including year t_0. For H_j, $j = 1, \ldots, m$, the first order approximation $h_j(\vec{\alpha}_{t_j}^*)$ is

Table 3.8: A summary of the distinctions among L, $\hat{L}$, l, H_j, $\hat{H}_j$ and h_j.

L	A non-linear function of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$
$\hat{L}$	An approximation of L, obtained by setting $\kappa_i(t) = \kappa_i(t_0) + C_1(t_0)(t - t_0)$ for $i = 1, 2$ and $t > t_0$; $\hat{L}$ is a non-linear function of $\kappa_1(t_0)$ and $\kappa_2(t_0)$ (i.e., $\vec{\alpha}^*_{t_0}$)
l	An approximation of L based on a first-order Taylor's expansion around $\hat{L}$; l is a linear function of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$
H_j	A non-linear function of $\vec{\alpha}^*_{t_j}$
$\hat{H}_j$	An approximation of H_j, obtained by setting $\kappa_i(t_j) = \kappa_i(t_0) + C_1(t_0)(t_j - t_0)$ for $i = 1, 2$; $\hat{H}_j$ is a non-linear function of $\kappa_1(t_0)$ and $\kappa_2(t_0)$ (i.e., $\vec{\alpha}^*_{t_0}$)
h_j	An approximation of H_j based on a first-order Taylor's expansion around $\hat{H}_j$; h_j is a linear function of $\vec{\alpha}^*_{t_j}$

given by

$$H_j \approx h_j(\vec{\alpha}^*_{t_j}) = \hat{H}_j + \left(\frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}\right)' (\vec{\alpha}^*_{t_j} - \hat{\vec{\alpha}}^*_{t_j}), \tag{3.7}$$

where $\hat{H}_j$ is defined in equation (3.5). For brevity, we suppress the arguments of l and h_j in the rest of this chapter. Also, unless otherwise specified, the partial derivatives $\partial L/\partial \vec{\alpha}^*_i$ and $\partial H_j/\partial \vec{\alpha}^*_{t_j}$ are evaluated at $\vec{\alpha}^*_i = \hat{\vec{\alpha}}^*_i$ and $\vec{\alpha}^*_{t_j} = \hat{\vec{\alpha}}^*_{t_j}$, respectively. To facilitate exposition, in Table 3.8 we summarize the distinctions among L, $\hat{L}$, l, H_j, $\hat{H}_j$ and h_j.

The hedging strategy is obtained by minimizing the variance of $l - \sum_{i=1}^m N_i h_i$ as of the time when the hedge is established. The variance to be minimized can be expressed as

$$\begin{aligned}
\mathrm{Var}\left(l - \textstyle\sum_{i=1}^m N_i h_i\right) &= \mathrm{Var}\left(\sum_{i=t_0+1}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}^*_i}\right)' (\vec{\alpha}^*_i - \hat{\vec{\alpha}}^*_i) - \sum_{j=1}^m N_j \left(\frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}\right)' (\vec{\alpha}^*_{t_j} - \hat{\vec{\alpha}}^*_{t_j})\right) \\
&= \mathrm{Var}\left(\sum_{\substack{i=t_0+1 \\ i \neq t_1, t_2, \ldots, t_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}^*_i}\right)' (\vec{\alpha}^*_i - \hat{\vec{\alpha}}^*_i) + \sum_{j=1}^m \left(\frac{\partial L}{\partial \vec{\alpha}^*_{t_j}} - N_j \frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}\right)' (\vec{\alpha}^*_{t_j} - \hat{\vec{\alpha}}^*_{t_j})\right).
\end{aligned}$$

It is interesting to note that $\mathrm{Var}(l - \sum_{i=1}^m N_i h_i)$ can be written as the sum of three com-

ponents:

$$\mathrm{Var}\left(l - \sum_{i=1}^{m} N_i h_i\right) = V_1 + V_2 + V_3,$$

where

$$V_1 = \sum_{\substack{i,j=t_0+1 \\ i,j\neq t_1,t_2,\ldots,t_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*) \left(\frac{\partial L}{\partial \vec{\alpha}_j^*}\right)$$

represents the variance contributed from the hidden states that are not related to the hedging instruments,

$$V_2 = \sum_{i=1}^{m}\sum_{j=1}^{m} \left(\frac{\partial L}{\partial \vec{\alpha}_{t_i}^*} - N_i \frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*}\right)' \mathrm{Cov}(\vec{\alpha}_{t_i}^*, \vec{\alpha}_{t_j}^*) \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)$$

represents the variance contributed from the hidden states that are directly related to the hedging instruments, and

$$V_3 = 2 \sum_{\substack{i=t_0+1 \\ i\neq t_1,t_2,\ldots,t_m}}^{t_0+T} \sum_{j=1}^{m} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*}\right)' \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_{t_j}^*) \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}\right)$$

represents the variance contributed from the interaction between the states that are related and unrelated to the hedging instruments. Note that V_1 is free of N_j, $j = 1, \ldots, m$. We can interpret V_1 to mean the risk that cannot be hedged by the collection of m hedging instruments.

All partial derivatives involved in $\mathrm{Var}(l - \sum_{i=1}^{m} N_i h_i)$ can be computed analytically. First, we rewrite L in terms of $\vec{\alpha}_t^*$, $t = t_0 + 1, \ldots, t_0 + T$, as

$$L = \sum_{u=1}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \left(1 + \exp(M_{x_0,t}' \vec{\alpha}_t^*)\right)^{-1}, \tag{3.8}$$

where $M_{x_0,t} = (1, x_0 + t - t_0 - 1 - \bar{x})'$. The partial derivative of L with respect to $\vec{\alpha}_i^*$,

evaluated at $\vec{\alpha}_i = \hat{\vec{\alpha}}_i$, can then be computed as

$$\frac{\partial L}{\partial \vec{\alpha}_i^*} = M_{x_0,i} \frac{\exp(M'_{x_0,i}\hat{\vec{\alpha}}_i^*)}{1+\exp(M'_{x_0,i}\hat{\vec{\alpha}}_i^*)} \left(- \sum_{u=i-t_0}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \left(1+\exp(M'_{x_0,t}\hat{\vec{\alpha}}_t^*)\right)^{-1} \right).$$

Similarly, we can express H_j in terms of $\vec{\alpha}_{t_j}^*$ as

$$H_j = e^{-r(t_j-t_0)} \left(q^f_{x_j,t_j} - 1 + \left(1+\exp((M^{(H)}_{x_j,t_j})'\vec{\alpha}_{t_j}^*)\right)^{-1} \right),$$

where $M^{(H)}_{x_j,t_j} = (1, x_j - \bar{x})'$. The partial derivative of H_j with respect to $\vec{\alpha}_{t_j}^*$, evaluated at $\vec{\alpha}_{t_j} = \hat{\vec{\alpha}}_{t_j}$, can then be calculated as

$$\frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} = -e^{-r(t_j-t_0)} M^{(H)}_{x_j,t_j} \frac{\exp\left((M^{(H)}_{x_j,t_j})'\hat{\vec{\alpha}}_{t_j}^*\right)}{\left(1+\exp\left((M^{(H)}_{x_j,t_j})'\hat{\vec{\alpha}}_{t_j}^*\right)\right)^2}.$$

The expression for $\mathrm{Var}(l - \sum_{i=1}^m N_i h_i)$ also involves $\mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$, the covariance matrix of $\vec{\alpha}_i^*$ and $\vec{\alpha}_j^*$. To compute $\mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$, we first calculate the covariance matrix of $\vec{\alpha}_i$ and $\vec{\alpha}_j$ as

$$\Xi_{i,j} := \mathrm{Cov}(\vec{\alpha}_i, \vec{\alpha}_j) = \begin{cases} A^{|i-j|}(Q + AQA' + \cdots A^{j-(t_0+1)}Q(A^{j-(t_0+1)})') & , i > j \\ (Q + AQA' + \cdots A^{i-(t_0+1)}Q(A^{i-(t_0+1)})')(A^{|i-j|})' & , i < j \\ Q + AQA' + \cdots A^{j-(t_0+1)}Q(A^{j-(t_0+1)})' & , i = j \end{cases}. \quad (3.9)$$

Then $\Xi_{i,j}$ is decomposed into four block matrices as follows:

$$\Xi_{i,j} = \begin{pmatrix} \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*) & \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^{**}) \\ \mathrm{Cov}(\vec{\alpha}_i^{**}, \vec{\alpha}_j^*) & \mathrm{Cov}(\vec{\alpha}_i^{**}, \vec{\alpha}_j^{**}) \end{pmatrix}.$$

Finally, $\Xi^*_{i,j} = \mathrm{Cov}(\vec{\alpha}_i^*, \vec{\alpha}_j^*)$ can be obtained from the upper-left block matrix in $\Xi_{i,j}$.

To derive the hedging strategy, we first take partial derivative of $\mathrm{Var}(l - \sum_{j=1}^m N_j h_j)$

with respect to N_i for $i = 1, 2, \ldots, m$:

$$\begin{aligned}\frac{\partial \mathrm{Var}(l-\sum_{j=1}^m N_j^* h_j)}{\partial N_i} &= \frac{\partial V_2}{\partial N_i} + \frac{\partial V_3}{\partial N_i} \\ &= -2\sum_{j=1}^m (\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,t_j}^* (\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*}) - 2 \sum_{\substack{j=t_0+1 \\ j\neq t_1,t_2,\ldots,t_m}}^{t_0+T} (\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\ &= 2\sum_{j=1}^m N_j (\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,t_j}^* \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} - 2 \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*}.\end{aligned}$$

Then the optimal hedging strategy is obtained by setting the partial derivatives to zero; that is,

$$\sum_{j=1}^m N_j (\frac{\partial H_1}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,t_j}^* \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} = \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_i}^*})' \Xi_{t_i,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*}, \quad i = 1, \ldots, m.$$

This system of linear equations can be written in matrix form as

$$\begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*} \\ \vdots & \ddots & \vdots \\ (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} = \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\ \vdots \\ \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \end{pmatrix}, \quad (3.10)$$

As to be explained in Section 3.5.5, the square matrix on the left-hand-side of the equation above is always invertible. Hence, the values of $N_1, \ldots, N_m$ can be obtained readily as follows:

$$\begin{pmatrix} N_1 \\ \vdots \\ N_m \end{pmatrix} = \begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*} \\ \vdots & \ddots & \vdots \\ (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,t_1}^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & \cdots & (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,t_m}^* \frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' \Xi_{t_1,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\ \vdots \\ \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_m}{\partial \vec{\alpha}_{t_m}^*})' \Xi_{t_m,j}^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \end{pmatrix}. \quad (3.11)$$

It is noteworthy that the formula above contains $\Xi_{s,t}^*$ for $s, t = t_1, \ldots, t_m$. As such, the

resulting hedging strategy incorporates the static and dynamic correlations between the hidden states in the observation equation.

3.5.4 Delta and Delta-Nuga Hedging Methods as Special Cases

In the generalized state-space hedging approach, we treat L and H_j as explicit functions of the state vectors $\vec{\alpha}_t^*$ for $t > t_0$. The analytical minimization of variance is made possible by approximating L and H_j with a first order Taylor's expansion around all state vectors involved. In contrast, in the traditional delta and delta-nuga hedging methods, the hedging strategies are derived on the basis of the best estimates L and H_j, which depend exclusively on $\vec{\alpha}_{t_0}^* = (\kappa_1(t_0), \kappa_2(t_0))'$.

In what follows, we show that the traditional delta and delta-nuga hedging methods are indeed special cases of the proposed generalized state-space hedging method when the following linear relations hold for $t > t_0 + 1$:

$$\begin{cases} \vec{\alpha}_t^* &= \vec{\alpha}_{t_0+1}^* + (t - (t_0+1)) \times \vec{\alpha}_{t_0+1}^{**} \\ \vec{\alpha}_t^{**} &= \vec{\alpha}_{t_0+1}^{**} \end{cases}. \tag{3.12}$$

3.5.4.1 Connections with the Traditional Delta Hedging Method

Suppose that the linear relations specified by equation (3.12) hold. Assume further that $m = 2$ hedging instruments are used and that the vector of drift terms is constant; i.e., $\vec{\alpha}_t^{**} = \vec{\hat{\alpha}}_{t_0}^{**}$ for $t \geq t_0 + 1$. Then $\Xi_{i,j}^* = Q^*$ for $i, j = t_0 + 1, \ldots, (t_0 + T)$, where Q^* is defined in equation (3.4). It follows that equation (3.10) can be reduced to

$$\begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' Q^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' Q^* \frac{\partial H_2}{\partial \vec{\alpha}_{t_2}^*} \\ (\frac{\partial H_2}{\partial \vec{\alpha}_{t_2}^*})' Q^* \frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*} & (\frac{\partial H_2}{\partial \vec{\alpha}_{t_2}^*})' Q^* \frac{\partial H_2}{\partial \vec{\alpha}_{t_2}^*} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}^*})' Q^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \\ \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_2}{\partial \vec{\alpha}_{t_2}^*})' Q^* \frac{\partial L}{\partial \vec{\alpha}_j^*} \end{pmatrix}. \tag{3.13}$$

Also, we have

$$\begin{aligned}\sum_{i=t_0+1}^{t_0+T}\frac{\partial L}{\partial\vec{\alpha}_i^*} &= -\sum_{i=t_0+1}^{t_0+T}M_{x_0,i}\hat{q}_{x_0+i-t_0-1,i}\left(\sum_{u=i-t_0}^{T}e^{-ru}\prod_{t=t_0+1}^{t_0+u}\hat{p}_{x_0+t-t_0-1,t}\right)\\ &= -\sum_{u=1}^{T}e^{-ru}\prod_{t=t_0+1}^{t_0+u}\hat{p}_{x_0+t-t_0-1,t}\left(\sum_{t=t_0+1}^{t_0+u}M_{x_0,t}\hat{q}_{x_0+t-t_0-1,t}\right)\\ &= \frac{\partial\hat{L}}{\partial\vec{\alpha}_{t_0}^*}\end{aligned} \tag{3.14}$$

and

$$\frac{\partial H_j}{\partial\vec{\alpha}_{t_j}^*} = -e^{-r(t_j-t_0)}M_{x_j,t_j}^{(H)}\hat{q}_{x_j,t_j}\times\hat{p}_{x_j,t_j} = \frac{\partial\hat{H}_j}{\partial\vec{\alpha}_{t_0}^*} \tag{3.15}$$

for $j = 1, 2$.

Substituting equations (3.14) and (3.15) into equation (3.13), we immediately obtain

$$\begin{pmatrix}\left(\frac{\partial\hat{H}_1}{\partial\vec{\alpha}_{t_0}^*}\right)'\\ \left(\frac{\partial\hat{H}_2}{\partial\vec{\alpha}_{t_0}^*}\right)'\end{pmatrix}Q^*\begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\vec{\alpha}_{t_0}^*} & \frac{\partial\hat{H}_2}{\partial\vec{\alpha}_{t_0}^*}\end{pmatrix}\begin{pmatrix}N_1\\ N_2\end{pmatrix} = \begin{pmatrix}\left(\frac{\partial\hat{H}_1}{\partial\vec{\alpha}_{t_0}^*}\right)'\\ \left(\frac{\partial\hat{H}_2}{\partial\vec{\alpha}_{t_0}^*}\right)'\end{pmatrix}Q^*\frac{\partial\hat{L}}{\partial\vec{\alpha}_{t_0}^*},$$

or equivalently,

$$\begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)}\\ \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)}\end{pmatrix}Q^*\begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)}\\ \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)}\end{pmatrix}\begin{pmatrix}N_1\\ N_2\end{pmatrix} = \begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)}\\ \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)}\end{pmatrix}Q^*\begin{pmatrix}\frac{\partial\hat{L}}{\partial\kappa_1(t_0)}\\ \frac{\partial\hat{L}}{\partial\kappa_2(t_0)}\end{pmatrix}.$$

By definition, Q^* is positive definite and hence invertible. If the matrix

$$\begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)}\\ \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)}\end{pmatrix}$$

is also invertible, then we have

$$\begin{pmatrix}N_1\\ N_2\end{pmatrix} = \begin{pmatrix}\frac{\partial\hat{H}_1}{\partial\kappa_1(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_1(t_0)}\\ \frac{\partial\hat{H}_1}{\partial\kappa_2(t_0)} & \frac{\partial\hat{H}_2}{\partial\kappa_2(t_0)}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial\hat{L}}{\partial\kappa_1(t_0)}\\ \frac{\partial\hat{L}}{\partial\kappa_2(t_0)}\end{pmatrix},$$

which yields the same values of N_1 and N_2 as those implied by the traditional delta-hedging method. As Q^* is being canceled out in the derivation, the strategy developed by the delta hedging method does not incorporate any information concerning the variances

and covariances of the hidden states.

3.5.4.2 Connections with the Traditional Delta-Nuga Hedging Method

Suppose that the linear relations specified by equation (3.12) hold and that $m = 4$ hedging instruments are used. Equation (3.10) then becomes

$$\begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}})' \Xi^*_{t_1,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}})' \Xi^*_{t_1,t_m} \frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}} \\ \vdots & \ddots & \vdots \\ (\frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}})' \Xi^*_{t_m,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & (\frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}})' \Xi^*_{t_m,t_m} \frac{\partial H_k}{\partial \vec{\alpha}^*_{t_m}} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}})' \Xi^*_{t_1,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \\ \vdots \\ \sum_{j=t_0+1}^{t_0+T} (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_4}})' \Xi^*_{t_4,j} \frac{\partial L}{\partial \vec{\alpha}^*_j} \end{pmatrix} \tag{3.16}$$

It can be shown easily that under the assumptions made, the covariance matrix $\Xi_{i,j}$ can be written as

$$\Xi_{i,j} = A^{i-(t_0+1)} Q (A^{j-(t_0+1)})'$$

for $i, j = t_0 + 1, \ldots, (t_0 + T)$. Consequently, we have

$$\begin{aligned} \left(\frac{\partial H_i}{\partial \vec{\alpha}^*_{t_i}}\right)' \Xi^*_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}} &= \begin{pmatrix} \frac{\partial H_i}{\partial \kappa_1(t_i)} & \frac{\partial H_i}{\partial \kappa_2(t_i)} & 0 & 0 \end{pmatrix} \Xi_{t_i,t_j} \begin{pmatrix} \frac{\partial H_j}{\partial \kappa_1(t_j)} & \frac{\partial H_j}{\partial \kappa_1(t_j)} & 0 & 0 \end{pmatrix}' \\ &= \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}}\right)' \Xi_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}} \\ &= \left(\frac{\partial H_i}{\partial \vec{\alpha}_{t_i}}\right)' A^{t_i-t_0} P (A^{t_j-t_0})' \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}}, \end{aligned}$$

for $i, j = 1, 2, \ldots, 4$, where $P = A^{-1}Q(A^{-1})'$. Also, we have

$$\begin{aligned}
(A^{t_j - t_0})' \tfrac{\partial H_j}{\partial \vec{\alpha}_{t_j}} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ t_j - t_0 & 0 & 1 & 0 \\ 0 & t_j - t_0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial H_j}{\partial \kappa_1(t_j)} \\ \frac{\partial H_j}{\partial \kappa_2(t_j)} \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ x_j - \bar{x} \\ t_j - t_0 \\ (t_j - t_0)(x_j - \bar{x}) \end{pmatrix} \left(-e^{-r(t_j - t_0)} \hat{p}_{x_j, t_j} \hat{q}_{x_j, t_j}\right) \\
&= \begin{pmatrix} \frac{\partial \hat{H}_j}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_j}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_j}{\partial C_1(t_0)} & \frac{\partial \hat{H}_j}{\partial C_2(t_0)} \end{pmatrix}' = \tfrac{\partial \hat{H}_j}{\partial \vec{\alpha}_{t_0}},
\end{aligned}$$

for $j = 1, 2, 3, 4$. Hence, the left-hand-side of equation (3.16) can be reduced to

$$\begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}})' A^{t_1 - t_0} \\ \vdots \\ (\frac{\partial H_4}{\partial \vec{\alpha}_{t_4}})' A^{t_4 - t_0} \end{pmatrix} P \left((\tfrac{\partial H_1}{\partial \vec{\alpha}_{t_1}})' A^{t_1 - t_0} \cdots (\tfrac{\partial H_4}{\partial \vec{\alpha}_{t_4}})' A^{t_4 - t_0} \right) = \begin{pmatrix} (\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}})' \\ \vdots \\ (\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}})' \end{pmatrix} P \left((\tfrac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}})' \cdots (\tfrac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}})' \right).$$

Furthermore, from equation (3.8) we obtain

$$\begin{aligned}
\sum_{i=t_0+1}^{t_0+T} (i - t_0) \tfrac{\partial L}{\partial \vec{\alpha}_i^*} &= \sum_{i=t_0+1}^{t_0+T} (i - t_0) M_{x_0, i} \hat{q}_{x_0 + i - t_0 - 1, i} \left(- \sum_{u=i-t_0}^{T} e^{-ru} \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0 + t - t_0 - 1, t} \right) \\
&= \sum_{u=1}^{T} e^{-ru} (-1) \prod_{t=t_0+1}^{t_0+u} \hat{p}_{x_0 + t - t_0 - 1, t} \left(\sum_{t=t_0+1}^{t_0+u} (t - t_0) M_{x_0, t} \hat{q}_{x_0 + t - t_0 - 1, t} \right),
\end{aligned}$$

which gives

$$\sum_{i=t_0+1}^{t_0+T} (A^{i - t_0})' \frac{\partial L}{\partial \vec{\alpha}_i} = \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}}.$$

Therefore, the right-hand-side of equation (3.16) can be reduced to

$$\begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}_{t_1}})' A^{t_1 - t_0} \\ \vdots \\ (\frac{\partial H_4}{\partial \vec{\alpha}_{t_4}})' A^{t_4 - t_0} \end{pmatrix} P \sum_{j=t_0+1}^{t_0+T} (A^{j - t_0})' \frac{\partial L}{\partial \vec{\alpha}_j} = \begin{pmatrix} (\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}})' \\ \vdots \\ (\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}})' \end{pmatrix} P \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}}.$$

Finally, we can rewrite equation (3.16) as

$$\begin{pmatrix} (\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}})' \\ \vdots \\ (\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}})' \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}} & \cdots & \frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} (\frac{\partial \hat{H}_1}{\partial \vec{\alpha}_{t_0}})' \\ \vdots \\ (\frac{\partial \hat{H}_4}{\partial \vec{\alpha}_{t_0}})' \end{pmatrix} P \frac{\partial \hat{L}}{\partial \vec{\alpha}_{t_0}},$$

or equivalently

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_1}{\partial C_2(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} \begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_1}{\partial C_2(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix} P \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \frac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix}$$

By definition, Q is positive definite and hence $P = A^{-1}Q(A^{-1})'$ is invertible. If the other matrices in the equation above are invertible, then we immediately obtain

$$\begin{pmatrix} N_1 \\ \vdots \\ N_4 \end{pmatrix} = \begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial \kappa_1(t_0)} \\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{H}_1}{\partial C_2(t_0)} & \cdots & \frac{\partial \hat{H}_4}{\partial C_2(t_0)} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \hat{L}}{\partial \kappa_1(t_0)} \\ \vdots \\ \frac{\partial \hat{L}}{\partial C_2(t_0)} \end{pmatrix},$$

which yields N_1, N_2, N_3 and N_4 that are exactly the same as those implied by the traditional delta-nuga hedging method. The hedging strategy is free of Q^* and therefore incorporates no information concerning the variances and covariances of the hidden states.

3.5.4.3 Distinctions between the Proposed Method and the Traditional Methods

Although the delta and delta-nuga methods can be viewed as special cases of the generalized state-space method, they are fundamentally different from the generalized state-space method in several ways.

First, under the generalized state-space method, the hedging strategy is derived by optimizing a specific objective function: $\text{Var}(L - \sum_{j=1}^{m} N_j H_j)$. In contrast, the delta and delta-nega methods rely merely on sensitivity matching and involves no optimization.

Second, through $\Xi^*_{s,t}$, the generalized state-space method incorporates both static and dynamic correlations between different hidden states. Such correlations are not taken into account in the delta and delta-nega methods.

Third, instead of assuming the linear relation specified in equation (3.12), the generalized state-space method recognizes that L and H_j are functions of $\vec{\alpha}^*_{t_0+1}, \ldots, \vec{\alpha}^*_{t_0+T}$ and $\vec{\alpha}^*_{t_j}$, respectively. Even if the same model is assumed, the generalized state-space method would still be different (in terms of both the derivation and the resulting notional amounts) from the delta and delta-nuga methods.

Finally, in the generalized state-space method, the drifts themselves are state variables. However, in the delta-nuga hedging method, the drift term in the time-varying dynamic factor is regarded as a constant parameter which is recalibrated from time to time.

3.5.5 Comments on the Hedging Methods

3.5.5.1 Sub-optimality of the Traditional Methods

In the previous sub-section, we have shown that the traditional delta and delta-nuga hedging methods are special cases of the generalized state-space hedging method when the linear relations specified in equation (3.12) hold. Equivalently speaking, if the linear relations do not hold, then the notional amounts $N_1, \ldots, N_m$ computed using delta or delta-nuga hedging methods do not minimize

$$\text{Var}\left(L - \sum_{j=1}^{m} N_j H_j\right),$$

the variance of the present values of all cash flows associated with the hedged position. Therefore, in general, the hedging strategies derived from the traditional methods are suboptimal relative to those derived from the generalized state-space approach. The degree of sub-optimality is quantified in Section 3.6 where a numerical illustration is presented.

3.5.5.2 The Singularity Problem

Recall that in the traditional delta hedging method, the solution to N_1 and N_2 exists only if the following matrix is invertible:

$$\begin{pmatrix} \frac{\partial \hat{H}_1}{\partial \kappa_1(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_1(t_0)} \\ \frac{\partial \hat{H}_1}{\partial \kappa_2(t_0)} & \frac{\partial \hat{H}_2}{\partial \kappa_2(t_0)} \end{pmatrix} = \begin{pmatrix} -e^{-r(t_1-t_0)}\hat{p}_{x_1,t_1}\hat{q}_{x_1,t_1} & -e^{-r(t_2-t_0)}\hat{p}_{x_2,t_2}\hat{q}_{x_2,t_2} \\ -e^{-r(t_1-t_0)}(x_1-\bar{x})\hat{p}_{x_1,t_1}\hat{q}_{x_1,t_1} & -e^{-r(t_2-t_0)}(x_2-\bar{x})\hat{p}_{x_2,t_2}\hat{q}_{x_2,t_2} \end{pmatrix}.$$

If the references ages of both q-forwards are the same (i.e., $x_1 = x_2$), then the second row of the matrix would become perfectly linearly dependent on the first row and hence the

square matrix is not invertible. Note that this problem exists even though the reference years t_1 and t_2 of the q-forwards are different.

The same problem also applies to the traditional delta-nuga hedging method, under which the solution to $N_1, \ldots, N_4$ exists only if the square matrix in equation (3.6) is invertible. The first three rows of the square matrix in equation (3.6) are respectively given by

$$\begin{aligned}
&\text{Row 1:} \quad \frac{\partial \hat{H}_j}{\partial \kappa_1(t_0)} = -e^{-r(t_j - t_0)} \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}, && j = 1,2,3,4; \\
&\text{Row 2:} \quad \frac{\partial \hat{H}_j}{\partial \kappa_2(t_0)} = -e^{-r(t_j - t_0)} (x_j - \bar{x}) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}, && j = 1,2,3,4; \\
&\text{Row 3:} \quad \frac{\partial \hat{H}_j}{\partial C_1(t_0)} = -e^{-r(t_j - t_0)} (t_j - t_0) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}, && j = 1,2,3,4.
\end{aligned}$$

It is clear that when the references ages of the q-forwards are the same (i.e., $x_1 = x_2 = x_3 = x_4$), the first and second rows in the matrix are perfectly linearly dependent on each other, which means the inverse of the matrix does not exist. On top of that, the problem of singularity will also occur if the q-forwards are linked to the same cohort; that is, $t_j - x_j = c$ for $j = 1, \ldots, 4$, where c denotes the year of birth to which the q-forwards are linked. This is because in this case, the difference between rows 3 and 2 becomes

$$-e^{-r(t_j - t_0)} (c - t_0 + \bar{x}) \hat{p}_{x_j,t_j} \hat{q}_{x_j,t_j}, \quad j = 1,2,3,4,$$

which is perfectly linearly dependent on row 1. This fact suggests that although it seems natural to choose q-forwards that are linked to the same cohort (as the one associated with the annuity liability), such a choice is not desirable if the delta-nuga hedging method is used.

The singularity problem does not apply to the generalized state-space hedging methodology, provided that the state vectors beyond year t_0 are not deterministically related as specified in equation (3.12). To explain why, let us have a closer scrutiny of the square matrix in equation (3.10). Because

$$\text{Cov}(h_i, h_j) = \left(\frac{\partial H_i}{\partial \vec{\alpha}^*_{t_i}}\right)' \Xi^*_{t_i,t_j} \frac{\partial H_j}{\partial \vec{\alpha}^*_{t_j}}$$

for $i, j = 1, 2, \ldots, m$, the square matrix

$$\Sigma_h := \begin{pmatrix} (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}})' \Xi^*_{t_1,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & (\frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}})' \Xi^*_{t_1,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \\ \vdots & \ddots & \vdots \\ (\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}})' \Xi^*_{t_m,t_1} \frac{\partial H_1}{\partial \vec{\alpha}^*_{t_1}} & \cdots & (\frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}})' \Xi^*_{t_m,t_m} \frac{\partial H_m}{\partial \vec{\alpha}^*_{t_m}} \end{pmatrix}$$

can be viewed as the covariance matrix of $\vec{h} = (h_1, \ldots, h_m)'$. By the spectral theorem, the square matrix can be written as

$$\Sigma_h = U \Lambda U',$$

where U is an orthogonal matrix and Λ is a diagonal matrix containing the eigenvalues of Σ_h. It immediately follows that

$$U' \Sigma_h U = \Lambda.$$

We can view $U' \Sigma_h U$ as the covariance matrix of the random vector $U' \vec{h}$, of which each element is a linear combination of the elements in $\vec{h}$. From equation (3.7) and the fact that $t_1 < \ldots < t_m$, we can infer that there does not exist a linear combination of $h_1, \ldots, h_m$ such that the linear combination is non-random. Therefore, all diagonal elements in Λ must be straightly positive. With straightly positive eigenvalues, the invertibility of Σ_h is guaranteed.

3.5.5.3 The Hedging Instrument Selection Problem

In using the traditional delta or delta-nuga hedging method, the number of hedging instruments must be either 2 (for delta) or 4 (for delta-nuga). In contrast, the generalized state-space hedging method can be implemented with any number of hedging instruments, as the solution to equation (3.10) exists for any $m \geq 1$. In this sense, the generalized state-space hedging method may be considered as more adaptable to different stages of market development.

As the market grows, more hedging instruments will become available. For reasons such as transaction costs, a hedger may wish to use only a subset of (rather than all) instruments that are available in the market. One possible criterion for choosing the subset of instruments is the resulting variance of the present values of all cash flows involved. To formulate this method mathematically, we first define a subset $\mathbb{M}$ which contains m

element(s) out of a set of k elements:

$$\mathbb{M} = \{(a_1, a_2, \ldots, a_m) : a_i \in \{1, 2, \ldots, k\}, a_1 < a_2 < \ldots < a_m\},$$

where m and k represent the number of instruments that the hedger wishes to use and the total number of instruments available in the market, respectively. The collection of m instrument(s) selected can written as

$$\underset{S_m \in \mathbb{M}}{\arg\min} \left\{ \text{Var} \left(\sum_{\substack{i=t_0+1 \\ i \neq t_j, j \in S_m}}^{t_0+T} \left(\frac{\partial L}{\partial \vec{\alpha}_i^*} \right)' (\vec{\alpha}_i^* - \hat{\vec{\alpha}}_i^*) + \sum_{j \in S_m} \left(\frac{\partial L}{\partial \vec{\alpha}_{t_j}^*} - N_j \frac{\partial H_j}{\partial \vec{\alpha}_{t_j}^*} \right)' (\vec{\alpha}_{t_j}^* - \hat{\vec{\alpha}}_{t_j}^*) \right) \right\}.$$

3.6 Illustrating the Hedging Methods

3.6.1 Assumptions

In this illustration, the liability being hedged is a 30-year life annuity immediate (i.e., $T = 30$). The annuity is sold to a person aged $x_0 = 70$ at the end of year $t_0 = 2010$, and a longevity hedge is established at the same time as the annuity is sold. We assume that the annuitant is subject to exactly the same mortality as Canadian males.

The hedging instruments under consideration are q-forwards. It is assumed that at the time when the hedge is established, there are exactly $k = 4$ q-forwards available in the market. The reference ages and years of these q-forwards are summarized below:

j	Reference age x_j	Reference year t_j
1	76	2017
2	85	2026
3	92	2033
4	100	2040

The first three q-forwards are associated with the same cohort (with year-of-birth 1941), whereas the last q-forward is associated with the cohort born one year earlier. We intentionally avoid having all four q-forwards linked to the same cohort, because otherwise the traditional delta-nuga hedging method would fail due to the singularity problem. Note that

the maturities of the q-forwards are approximately $T/4$, $T/2$, $3T/4$ and T, respectively. We assume that the q-forwards are also linked to the mortality of Canadian males.

The procedure which we use to assess hedge effectiveness can be summarized as follows:

(i) Preparation

- Compute the partial derivatives involved.
- Assuming that the hedger wishes to use m q-forwards, select m out of the $k = 4$ available q-forwards on the basis of variance minimization.
- Calculate the optimal notional amounts.

(ii) Simulation

- Simulate 10,000 mortality scenarios from the CBD/LLCBD model that is fitted to the data for Canadian males over a calibration window of 1941-2010 and an age range of 50-89.
- For each simulated mortality scenario, calculate the realized values of L and H_j, $j = 1, \ldots, m$. An interest rate of $r = 0.01$ is used for discounting purposes.

(iii) Evaluation

- On the basis of the 10,000 realizations of L and H_j, $j = 1, \ldots, m$, compute the value HE.

The objectives of this illustration are threefold: (1) to compare the hedge effectiveness produced by the traditional and newly proposed methods; (2) to assess how the negligence of stochastic drifts may affect the resulting hedge effectiveness; (3) to demonstrate the interaction among the model assumption, the hedging method and the duration of the liability being hedged. To better achieve these objectives, we divide the empirical results into the following four groups, depending on the hedging method and mortality model from the hedging strategy is derived:

	The mortality model on which the hedging strategy is based	Hedging method
Group 1	The original CBD model	Traditional delta hedging
Group 2	The LLCBD model	Traditional delta and delta-nuga hedging
Group 3	The original CBD model	Generalized state-space hedging
Group 4	The LLCBD model	Generalized state-space hedging

Table 3.9: The hedge effectiveness and notional amounts for all possible combinations of $m = 2$ q-forwards, Groups 1 and 3. The '–' sign indicates that the corresponding q-forward is not used. The simulation model is the original CBD model.

Group 1 Assumed model: The original CBD model Method: Traditional delta hedging					Group 3 Assumed model: The original CBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
0.9230	98.0364	47.4528	–	–	0.9358	71.5625	47.9503	–	–
0.8997	137.3202	–	17.0804	–	0.9135	107.8487	–	18.1090	–
0.8183	154.1561	–	–	7.5688	0.8322	132.5633	–	–	7.0199
−0.3565	–	165.8758	−42.6257	–	0.9053	–	44.2286	8.4076	–
0.0177	–	130.3489	–	−13.2221	0.8924	–	51.7275	–	2.4632
−5.2683	–	–	156.3945	−61.7343	0.7463	–	–	24.6273	−0.2105

3.6.2 Result I: A Comparison of Different Hedging Methods

In this sub-section, we focus on comparing the hedge effectiveness produced by different hedging methods. The relevant results are displayed in Tables 3.9 and 3.10.

Let us first study Table 3.9, which compares the results in Groups 1 and 3. For these two groups, the hedging strategies are based on the same model (the original CBD model) but are derived using different hedging techniques. Here, we intend to focus on the difference in hedging techniques, so we assume that the actual (simulation) mortality model is also the original CBD model. We report the results when $m = 2$ hedging instruments are used, because the results in Group 1 are derived from the traditional delta-hedging method which requires exactly two hedging instruments. For Group 1, the best hedge effectiveness obtained is greater than 92%, but different combinations of q-forwards lead to highly different values of HE. In the worst case when the third and fourth q-forwards are used, the value of HE is even negative, which means the seller of the annuity is subject to even more longevity risk when the hedge is in place. For Group 3, the values of HE are much less sensitive to the choice of the two q-forwards. Also, given the same choice of q-forwards, the hedge effectiveness in Group 3 is always higher than that in Group 1, indicating that hedgers can make better use of the hedging instruments if they use the generalized state-space hedging method. This advantage may be attributed to the fact that the generalized state-space hedging method incorporates information about the static and dynamic correlations between the hidden states, but the delta hedging method does not.

We then move on to studying Table 3.10, which compares the results from Groups 2 and 4. As before, the hedging strategies for these two groups are based on the same model (the

Table 3.10: The hedge effectiveness and notional amounts for all possible combinations of $m = 2$ and $m = 4$ q-forwards, Groups 2 and 4. The '–' sign indicates that the corresponding q-forward is not used. The simulation model is the LLCBD model.

Group 2 Assumed model: The LLCBD model Method: Traditional delta/delta-nuga hedging					Group 4 Assumed model: The LLCBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-forwards used: $m = 2$									
0.7209	110.4838	68.2613	–	–	0.8409	−25.5044	76.0989	–	–
0.8701	158.4984	–	27.8511	–	0.8957	97.7825	–	24.9453	–
0.1140	179.0761	–	–	13.2179	0.6569	92.8728	–	–	7.4087
−3.5174	–	225.3339	−64.0867	–	0.9579	–	41.0866	13.4239	–
−5.7736	–	178.2121	–	−21.2905	0.9386	–	56.9217	–	3.3286
−32.7169	–	–	242.3726	−101.8099	0.8764	–	–	30.2868	−2.9065
Number of q-forwards used: $m = 4$									
0.8754	157.7860	1.0128	27.4379	−0.0000	0.9737	66.7273	38.9987	13.2875	0.8685

LLCBD model) but different hedging techniques. For Group 4, all results are derived from the generalized state-space hedging method, whereas for Group 2, the results for $m = 2$ and $m = 4$ are respectively obtained by using the traditional delta and delta-nuga hedging methods. The simulation model used here is the LLCBD model, consistent with the model on which the hedging strategies are based. Compared to the values of HE in Group 2, the values of HE in Group 4 are consistently higher and are more robust relative to the choice of hedging instruments. It is also noteworthy that some values of HE in Group 2 are negative, as the traditional hedging methods do not guarantee a reduction in variance. This problem is discussed further in Section 3.6.5.1.

The results presented in this sub-section demonstrate the previously made argument concerning the sub-optimality of the traditional hedging methods. Other things equal, the generalized state-space hedging method yields a better hedge effectiveness, regardless of whether the original CBD model or the LLCBD model is assumed.

In this sub-section, the values of HE for Groups 1 and 3 are calculated using scenarios generated from the original CBD model, while those for Groups 2 and 4 are computed using scenarios simulated from the LLCBD model. For the readers' information, the estimated values of $\text{Var}(L)$ under the CBD and LLCBD models are 0.0913 and 0.0715, respectively. The reason that $\text{Var}(L)$ calculated using the CBD scenarios is larger can be understood by revisiting Figure 3.7, which shows that the LLCBD fan charts are initially narrower than the corresponding CBD fan charts, but become wider about 20 years from the forecast origin as they widen at faster rates. The reason behind can also be visualized in Figure

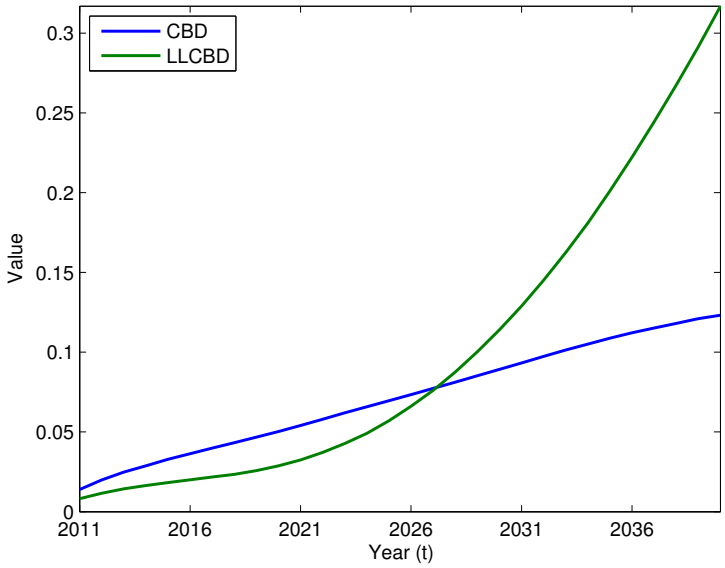


Figure 3.15: The standard deviations of the annuitants' cohort death probabilities in logit scale (i.e., $\ln(q_{t-1941,t}/(1-q_{t-1941,t}))$ for $t = 2011, \ldots, 2041$), estimated using the CBD and LLCBD models.

3.15, which compares the standard deviations of the cohort death probabilities of the annuitants (in logit scale) under the two models. The standard deviations under the CBD model are higher for the first half of the annuity's maximum duration, but the opposite is true for the second half. Because cash flows in distant future are less influential (due to the effects of discounting and survivorship), $\text{Var}(L)$ depends more heavily on the uncertainty surrounding the earlier cohort death probabilities and is consequently higher under the CBD model.

3.6.3 Result II: The Impact of Model Mis-Specification

We have previously argued that compared to the original CBD model, the LLCBD model (with stochastic drifts) may more realistically represent the true underlying mortality dynamics. In this sub-section, we examine how much hedge effectiveness may be lost if the true underlying mortality dynamics are driven by a model with stochastic drifts while a model with constant drifts is assumed in deriving the hedging strategies. To achieve this goal, we compare the results in Groups 3 and 4 (see Table 3.11) when the true (simulation) model is the LLCBD model.

Table 3.11: The hedge effectiveness and the corresponding notional amounts when $m = 1, 2, 3, 4$ q-forwards are used, Groups 3 and 4. The '–' sign indicates that the corresponding q-forward is not used. The simulation model is the LLCBD model.

Group 3 Assumed model: The original CBD model Method: Generalized state-space hedging					Group 4 Assumed model: The LLCBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-forwards used: $m = 1$									
0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
Number of q-forwards used: $m = 2$									
0.6496	71.5625	47.9503	–	–	0.9579	–	41.0866	13.4239	–
Number of q-forwards used: $m = 3$									
0.8557	75.7070	29.6824	9.3110	–	0.9732	64.1560	36.5098	16.1386	–
Number of q-forwards used: $m = 4$									
0.8728	76.1325	29.7784	7.9015	0.7337	0.9737	66.7273	38.9987	13.2875	0.8685

All results in Groups 3 and 4 are based on the generalized state-space hedging method and mortality scenarios simulated from the LLCBD model. The differences between the results in these two groups are because of the difference in the model from which the notional amounts are derived. Since the generalized state-space hedging method permits us to use any number of hedging instruments, we consider both the situation when all available q-forwards are used and the situation when only a subset of the available q-forwards are included. When less than four q-forwards are used, the q-forwards are chosen using the method described in Section 3.5.5.3. For Group 3, HE does not necessarily increase with m, because the model on which the hedging strategies are based is inconsistent with the simulation model.

Compared to the corresponding values in Group 3, the values of HE in Group 4 are consistently higher. This observation suggests that the negligence of stochastic drifts may result in a material loss of hedge effectiveness, if the true underlying model is one with stochastic drifts.

3.6.4 Result III: The Interaction among Different Factors

In this sub-section, we demonstrate the interaction among the model assumption, the hedging method and the duration of the liability being hedged. To achieve this goal, we now consider annuity liabilities with durations ranging from $T = 25$ to $T = 40$ years.

For each annuity liability, the longevity hedge is composed of $m = 2$ q-forwards, which are linked to the same cohort of individuals as the annuity liability and have maturities of (approximately) $T/2$ and $T/4$. All other previously made assumptions, including the assumptions about the values of x_0, t_0 and r, remain unchanged. The simulation model used is still the LLCBD model. The results are presented in Figure 3.16.

We first compare the trends of HE for Groups 2 and 4. As expected, the trend for Group 4, which is based on the generalized state-space hedging method, is always higher than that for Group 2, which is based on the traditional delta hedging method. A more interesting observation is that the gap between the two trends is roughly constant. This observation indicates that the benefit from using the more general hedging method is somewhat fixed, with little dependence on the duration of the liability being hedged.

Next, we compare the trends of HE for Groups 3 and 4. The gap between the two trends widens rapidly as the duration of the liability being hedged increases. From this observation we can infer that when the true model is one with stochastic drifts, the benefit of incorporating stochastic drifts into the hedging strategy is more remarkable if the liability being hedged is longer-dated. This conclusion is reasonable because the assumption about the drifts, which determine the gradients of future mortality trends, should have more long-run than short-run effects.

Finally, we compare the trends of HE for Groups 1 and 4. The gap between the HE trends for these two groups is approximately the sum of the gap between the HE trends for Groups 2 and 4 and the gap between the HE trends for Groups 3 and 4. It reflects the overall benefit of using both the generalized state-space hedging method and the assumption of stochastic drifts in deriving the hedging strategies.

3.6.5 Further Issues

3.6.5.1 Why Delta and Delta-Nuga Methods May Perform Unsatisfactorily Sometimes?

In Section 3.6.2 we reveal that the delta and delta-nuga hedging methods may sometimes lead to a low or even negative hedge effectiveness. We may attribute this problem to the linearity assumption on which these methods are based.

To illustrate, we construct a scenario analysis that is based on the following assumptions:

- Mortality model on which the hedging strategy is based: the original CBD model

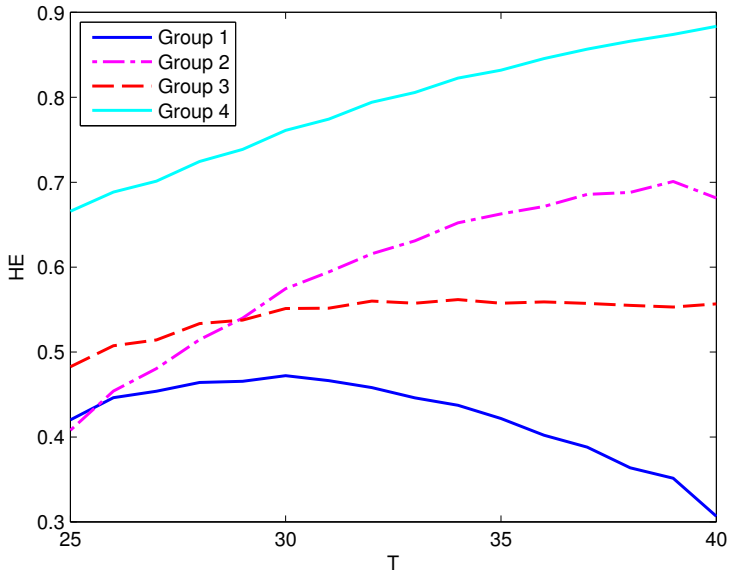


Figure 3.16: The relationship between the hedge effectiveness (HE) and the duration (T) of the liability being hedged, Groups 1, 2, 3 and 4. The simulation model is the LLCBD model.

- Simulation model: the original CBD model
- Hedging instruments: the 2nd and 3rd q-forward, $(x_2, t_2) = (85, 2026)$ and $(x_3, t_3) = (92, 2033)$.

Under these assumptions, the values of HE produced by the delta hedging method and the generalized state-space hedging method are -0.3565 and 0.9053, respectively (see Table 3.9).

Six mortality scenarios – built on hypothetical sample paths of $\kappa_1(t)$ and $\kappa_2(t)$ – are used to analyze why the two methods perform so differently (see Figure 3.17). Scenarios (i) and (ii) are the most extreme 'linear' scenarios within the 95% prediction intervals, while Scenarios (iii) to (vi) are extreme 'non-linear' scenarios within the 95% prediction intervals.[9] For each scenario and hedging method, we calculate the realized values of

[9]The annuity liability and all hedging instruments used are related to ages greater than $\bar{x} = 69.5$, so the projected mortality would be high (low) when both $\kappa_1(t)$ and $\kappa_2(t)$ are high (low). It follows that (1) and (a) are the linear sample paths that would result in the highest projected mortality, whereas (2) and (b) are the linear sample paths that would lead to the lowest projected mortality.

Table 3.12: The values of $L - N_2H_2 - N_3H_3 - \hat{L}$ and $L - \hat{L}$ under the six hypothetical extreme mortality scenarios described in Figure 3.17.

Scenario	Sample paths	N_2H_2	N_3H_3	$L - N_2H_2 - N_3H_3 - \hat{L}$	$L - \hat{L}$
Delta Hedge					
(i)	(1) × (a)	−1.3804	0.9932	−0.1379	−0.5251
(ii)	(2) × (b)	1.2271	−0.8361	0.1658	0.5568
(iii)	(3) × (c)	−1.9107	−0.9401	2.5764	−0.2743
(iv)	(3) × (d)	−0.0769	0.1562	−0.1848	−0.1055
(v)	(4) × (c)	0.0511	−0.1431	0.1883	0.0963
(vi)	(4) × (d)	1.6191	1.1475	-2.5397	0.2270
Generalized State-Space Hedge					
(i)	(1) × (a)	−0.3681	−0.1959	0.0389	−0.5251
(ii)	(2) × (b)	0.3272	0.1649	0.0647	0.5568
(iii)	(3) × (c)	−0.5095	0.1854	0.0497	−0.2743
(iv)	(3) × (d)	−0.0205	−0.0308	−0.0542	−0.1055
(v)	(4) × (c)	0.0136	0.0282	0.0545	0.0963
(vi)	(4) × (d)	0.4317	−0.2263	0.0216	0.2270

$L - N_2H_2 - N_3H_3 - \hat{L}$ (the hedged position) and $L - \hat{L}$ (the unhedged position). The results are tabulated in Table 3.12.

For Scenarios (i) and (ii), the hedge derived using the delta hedging method yields values of $L - N_2H_2 - N_3H_3 - \hat{L}$ that are much smaller than $L - \hat{L}$ in magnitude, suggesting that the hedge can withstand extreme mortality scenarios that are 'linear'. However, for remaining four scenarios, all of which are 'non-linear', the hedge derived using the delta hedging method performed very badly (even worse than not hedging), indicating that the hedge is vulnerable to non-linear changes in the dynamic factors. We also observe that $L - N_2H_2 - N_3H_3 - \hat{L}$ and $L - \hat{L}$ may take the same or different signs, which means that the poor performance is sometimes because the hedge has gone in the wrong direction and sometimes because the hedge has gone in the right direction but too far.

For all six scenarios, the generalized state-space hedging method performs satisfactorily, yielding values of $L - N_2H_2 - N_3H_3 - \hat{L}$ that are much smaller than $L - \hat{L}$ in magnitude.

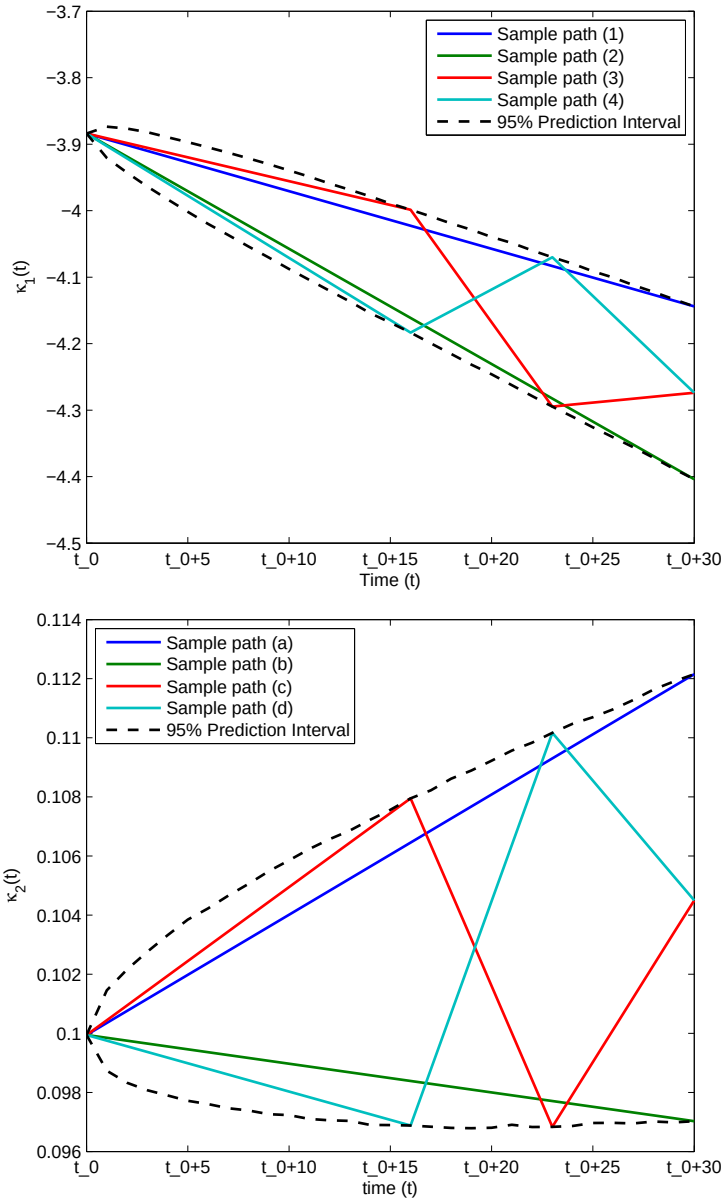


Figure 3.17: Six extreme mortality scenarios: Scenarios (i) to (vi) are formed by (1) $\times$ (a), (2) $\times$ (b), (3) $\times$ (c), (3) $\times$ (d), (4) $\times$ (c), (4) $\times$ (d), respectively. The dotted lines represent the 95% prediction intervals.

Table 3.13: The 'best achievable' hedges given the 10,000 mortality scenarios simulated from (a) the original CBD model and (b) the LLCBD model.

(a) Simulation model: the original CBD model				
HE	N_1	N_2	N_3	N_4
0.9358	−70.9784	−47.8004	–	–
0.9136	−107.4581	–	−17.8760	–
0.8322	−132.2430	–	–	−6.9878
0.9054	–	−44.4610	−8.0472	–
0.8926	–	−52.0698	–	−2.2432
0.7465	–	–	-24.3666	0.2981
(b) Simulation model: the LLCBD model				
HE	N_1	N_2	N_3	N_4
0.8410	−22.8565	75.3767	–	–
0.8967	93.0628	–	24.1189	–
0.6609	79.1050	–	–	6.8454
0.9586	–	41.4408	12.7478	–
0.9400	–	57.0345	–	3.0335
0.8769	–	–	29.1619	−2.6613
0.9749	62.2898	39.3229	13.0727	0.6231

3.6.5.2 The Best Achievable Hedge Effectiveness

Given the 10,000 mortality scenarios simulated from the CBD/LLCBD model, we can iteratively search for the combination of $N_1, \ldots, N_m$ what would maximize HE. The resulting value of HE may be regarded as the 'best achievable' hedge effectiveness (given the 10,000 scenarios) and can be used as a benchmark to assess the quality of the proposed hedging strategy.

We consider hedge portfolios with $m = 2, 4$ q-forwards, chosen from the collection of four q-forwards described in Section 3.6.1. Table 3.13 displays, for each portfolio, the best achievable value of HE and the corresponding values of $N_1, \ldots, N_m$ on the basis of the mortality scenarios simulated from (a) the original CBD model and (b) the LLCBD model. The values reported in Table 3.13 (a) are very close to those shown in Table 3.9 (Group 3), whereas the values reported in Table 3.13 (b) are very close to those shown in Table 3.10 (Group 4). The results suggest that the performance of the generalized state-space hedging strategy is nearly as good as the best achievable one.

3.6.5.3 Poisson Risk

The baseline results assume no Poisson risk (a.k.a. sampling risk). We now incorporate Poisson uncertainty into the hedging results by treating the cohort of annuitants as a random survivorship group.

Let $l_{x,t}$ be the number of annuitants who survive to age x at the beginning of year t, and $d_{x,t}$ be the number of annuitants who die in year t (between age x and $x+1$). For a fixed initial number of annuitants l_{x_0,t_0+1}, we have the following:

$$d_{x_0+t-t_0-1,t} \sim \text{Poisson}(l_{x_0+t-t_0-1,t} \times \tilde{q}_{x_0+t-t_0-1,t})$$

and

$$l_{x_0+t-t_0,t+1} = l_{x_0+t-t_0-1,t} - d_{x_0+t-t_0-1,t},$$

where $t = t_0+1, \ldots, t_0+T$ and $\tilde{q}_{x,t}$ is the underlying unobserved probability of death in year t (between age x and $x+1$).

In generating the hedging results, we simulate, for each of the 10,000 simulated mortality scenarios from the LLCBD model, one realization of $\{l_{x_0+t-t_0+1,t+1}; t = t_0, t_0+1, \ldots, t_0+T-1\}$. The per contract value of the liability being hedged in each mortality scenario is then computed as follows:

$$L = \sum_{u=1}^{T} e^{-ru} \frac{l_{x_0+u,t_0+1+u}}{l_{x_0,t_0+1}}.$$

The calculation of $N_1, \ldots, N_m$ remains the same as when Poisson uncertainty is not taken into account, because the q-forward hedge aims to mitigate only systematic longevity risk. There is also no change to the calculation of H_j, $j = 1, \ldots, m$, because the q-forwards are assumed to be linked to the smoothed death probabilities that are free of Poisson uncertainty.

We calculate the hedge effectiveness for annuity liabilities with different initial numbers of annuitants: $l_{x_0,t_0+1} = 10^3, 10^4, 10^5, 10^6, 10^7$. The results (see Table 3.14) are based on hedge portfolios with $m = 1, 2, 3, 4$ q-forwards, built by applying the generalized state-space method to the original CBD model (Group 3) and the LLCBD model (Group 4). When there are only 1 000 annuitants initially, Poisson risk erodes hedge effectiveness by about 30 to 40 percentage points. The erosion in hedge effectiveness is much milder (less than 10 percentage points) when the initial number of annuitants is 10 000, and becomes negligible when the initial number of annuitant grows to 100 000. When the initial number

of annuitants is 10^7, the resulting values of HE are identical (to four significant figures) to those when Poisson risk is assumed to be absent. Our results are in line with those produced by Li and Hardy (2011) and Cairns et al. (2014).

3.6.5.4 Sensitivity to the Covariance between State Vectors

Recall that Q represents the covariance matrix of the innovation vectors, thereby governing the static and dynamic correlations between the hidden states. The hedging strategy derived from the generalized state-space method depends on Q, but those derived from the delta- and delta-nuga methods do not. Hence, it is warranted to examine how the performance of different hedging strategies may change when Q is altered.

We consider the following five hypothetical situations:

(a) all elements are scaled up or down;

(b) elements related to $\kappa_1(t)$ (i.e., $Q_{1,i}$ and $Q_{i,1}$ for $i = 1, 2, 3, 4$) are scaled up or down;

(c) elements related to $\kappa_2(t)$ (i.e., $Q_{2,i}$ and $Q_{i,2}$ for $i = 1, 2, 3, 4$) are scaled up or down;

(d) elements related to $C_1(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are scaled up or down;

(e) elements related to $C_2(t)$ (i.e., $Q_{4,i}$ and $Q_{i,4}$ for $i = 1, 2, 3, 4$) are scaled up or down.

Four scaling factors, 0.2, 0.5, 2 and 5, are considered. The simulation model used is the LLCBD model (with the altered covariance matrix). The test results for hedges with $m = 4$ q-forwards are presented in Table 3.15.

As expected, when the delta-nuga hedging method is used, altering Q has no impact on the values of $N_1, \ldots, N_4$. However, the value of HE changes considerably when Q is scaled up or down. The value of HE is the lowest (0.1753) when the elements related to $C_2(t)$ are scaled up by a factor of 5.

When the generalized state-space hedging method is used, the values of $N_1, \ldots, N_4$ are adaptive to the modifications made to Q. As the notional amounts are 'corrected' accordingly[10], the resulting value of HE is much more robust relative to changes in Q. For every situation under consideration, the value of HE produced by the generalized state-space hedging method is higher than that produced by the delta-nuga hedging method.

[10]Situation (a) is an exception. If all elements of Q are scaled by the same factor, then according to equation (3.9) all elements in $\Xi^*_{i,j}$, for any i and j, would also be scaled by exactly the same factor. As a result, we can cancel out the scaling factor in equation (3.11), resulting in no change in the values of $N_1, \ldots, N_4$.

Table 3.14: The calculated values of HE when Poisson risk is absent and present. The hedge portfolios are composed of $m = 1, 2, 3, 4$ q-forwards and are built by applying the GSS method to the original CBD model (Group 3) and the LLCBD model (Group 4).

Without Poisson Risk										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
$m = 2$	0.6496	71.5625	47.9503	–	–	0.9579	–	41.0866	13.4239	–
$m = 3$	0.8557	75.7070	29.6824	9.3110	–	0.9732	64.1560	36.5098	16.1386	–
$m = 4$	0.8728	76.1325	29.7784	7.9015	0.7337	0.9737	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 1,000$										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.4271	–	60.0893	–	–	0.4487	–	–	22.3178	–
$m = 2$	0.3451	71.5625	47.9503	–	–	0.5062	–	41.0866	13.4239	–
$m = 3$	0.4528	75.7070	29.6824	9.3110	–	0.5143	64.1560	36.5098	16.1386	–
$m = 4$	0.4618	76.1325	29.7784	7.9015	0.7337	0.5151	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 10,000$										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7358	–	60.0893	–	–	0.7829	–	–	22.3178	–
$m = 2$	0.5986	71.5625	47.9503	–	–	0.8776	–	41.0866	13.4239	–
$m = 3$	0.7876	75.7070	29.6824	9.3110	–	0.8921	64.1560	36.5098	16.1386	–
$m = 4$	0.8027	76.1325	29.7784	7.9015	0.7337	0.8926	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 100,000$										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7922	–	60.0893	–	–	0.8503	–	–	22.3178	–
$m = 2$	0.6432	71.5625	47.9503	–	–	0.9499	–	41.0866	13.4239	–
$m = 3$	0.8475	75.7070	29.6824	9.3110	–	0.9648	64.1560	36.5098	16.1386	–
$m = 4$	0.8644	76.1325	29.7784	7.9015	0.7337	0.9653	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 1,000,000$										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.7994	–	60.0893	–	–	0.8567	–	–	22.3178	–
$m = 2$	0.6491	71.5625	47.9503	–	–	0.9570	–	41.0866	13.4239	–
$m = 3$	0.8550	75.7070	29.6824	9.3110	–	0.9723	64.1560	36.5098	16.1386	–
$m = 4$	0.8720	76.1325	29.7784	7.9015	0.7337	0.9727	66.7273	38.9987	13.2875	0.8685
With Poisson Risk, $l_{x_0,t_0+1} = 10,000,000$										
			Group 3					Group 4		
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
$m = 1$	0.8000	–	60.0893	–	–	0.8575	–	–	22.3178	–
$m = 2$	0.6496	71.5625	47.9503	–	–	0.9578	–	41.0866	13.4239	–
$m = 3$	0.8557	75.7070	29.6824	9.3110	–	0.9731	64.1560	36.5098	16.1386	–
$m = 4$	0.8727	76.1325	29.7784	7.9015	0.7337	0.9736	66.7273	38.9987	13.2875	0.8685

Table 3.15: The values of HE, N_1, N_2, N_3 and N_4 when Q is altered in different manners.

Multiplier	Group 2 Assumed model: The LLCBD model Method: Traditional delta-nuga hedging					Group 4 Assumed model: The LLCBD model Method: Generalized state-space hedging				
	HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
(a) All elements are scaled up or down										
0.2	0.8903	157.7860	1.0128	27.4379	-0.0000	0.9803	66.7273	38.9987	13.2875	0.8685
0.5	0.8849	157.7860	1.0128	27.4379	-0.0000	0.9779	66.7273	38.9987	13.2875	0.8685
2	0.8544	157.7860	1.0128	27.4379	-0.0000	0.9641	66.7273	38.9987	13.2875	0.8685
5	0.7788	157.7860	1.0128	27.4379	-0.0000	0.9287	66.7273	38.9987	13.2875	0.8685
(b) Elements related to $\kappa_1(t)$ (i.e., $Q_{1,i}$ and $Q_{i,1}$ for $i = 1, 2, 3, 4$) are scaled up or down										
0.2	0.9583	157.7860	1.0128	27.4379	-0.0000	0.9867	87.1269	41.1404	13.2163	1.1126
0.5	0.9401	157.7860	1.0128	27.4379	-0.0000	0.9838	78.9923	40.7965	13.2949	1.0401
2	0.3661	157.7860	1.0128	27.4379	-0.0000	0.9002	50.0976	33.4547	12.4363	0.1343
5	0.4538	157.7860	1.0128	27.4379	-0.0000	0.8918	62.6005	32.7750	11.8342	-2.9903
(c) Elements related to $\kappa_2(t)$ (i.e., $Q_{2,i}$ and $Q_{i,2}$ for $i = 1, 2, 3, 4$) are scaled up or down										
0.2	0.9306	157.7860	1.0128	27.4379	-0.0000	0.9821	67.7555	40.5623	13.2294	0.9485
0.5	0.9151	157.7860	1.0128	27.4379	-0.0000	0.9801	67.4009	39.9414	13.2894	0.9192
2	0.7206	157.7860	1.0128	27.4379	-0.0000	0.9450	65.5549	37.4718	13.1057	0.6844
5	0.6440	157.7860	1.0128	27.4379	-0.0000	0.9133	78.7940	37.4970	12.9966	0.4223
(d) Elements related to $C_1(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are scaled up or down										
0.2	0.8134	157.7860	1.0128	27.4379	-0.0000	0.9673	48.8085	33.5628	12.5653	1.0407
0.5	0.7978	157.7860	1.0128	27.4379	-0.0000	0.9745	21.5914	30.1378	12.2326	0.8986
2	0.9412	157.7860	1.0128	27.4379	-0.0000	0.9781	82.4635	41.9468	13.3386	0.9666
5	0.8541	157.7860	1.0128	27.4379	-0.0000	0.8821	85.7639	43.2876	12.6638	1.1792
(e) Elements related to $C_2(t)$ (i.e., $Q_{3,i}$ and $Q_{i,3}$ for $i = 1, 2, 3, 4$) are scaled up or down										
0.2	0.9161	157.7860	1.0128	27.4379	-0.0000	0.9784	84.0152	40.8037	13.4928	1.0814
0.5	0.8608	157.7860	1.0128	27.4379	-0.0000	0.9684	79.1166	40.3986	13.6192	0.7173
2	0.8781	157.7860	1.0128	27.4379	-0.0000	0.9480	53.2548	39.4064	13.2301	1.0908
5	0.1753	157.7860	1.0128	27.4379	-0.0000	0.5724	74.0987	41.8060	12.9891	1.1607

3.6.5.5 Sensitivity to the Number of q-forwards Used

As previously mentioned, the generalized state-space hedging method can be implemented with any number of q-forwards (m), because the solution to equation (3.10) exists for any $m \geq 1$. We now examine how the performance of the generalized state-space hedging method may change as the number of q-forwards used increases.

We consider the following collection of q-forwards:

j	Reference age x_j	Reference year t_j	j	Reference age x_j	Reference year t_j
1	80	2025	6	80	2030
2	80	2026	7	80	2031
3	80	2027	8	80	2032
4	80	2028	9	80	2033
5	80	2029	10	80	2034

Note that it is not possible to use this collection of q-forwards to build delta or delta-nuga hedges, because they are all linked to the same reference age.[11]

When $m = 1$, the hedge contains q-forward $j = 1$; when $m = 2$, the hedge contains q-forwards $j = 1, 2$; and so on. The hedging results for $m = 1, \ldots, 10$ are reported in Table 3.16. As m increases, the value of HE increases. However, the rate of increase in HE reduces with m, suggesting that the marginal benefit of adding an additional q-forward becomes small when the hedge portfolio already contains a large number of q-forwards.

When m becomes very large, inverting Σ_h may become difficult as it may be close to singular due to its large dimension. Users of this method should be mindful of this potential problem, although having a very large m is not likely.

3.6.5.6 Sensitivity to Choice of q-forwards

Finally, we examine the sensitivity of the hedging results to the choice of q-forwards. The following set of q-forwards is considered:

j	Reference age x_j	Reference year t_j
1	70	2020
2	88	2025
3	75	2030
4	77	2035

[11] When the q-forwards are linked to the same reference age, the delta and delta-nuga methods are subject to the singularity problem (see Section 3.5.5.2).

Table 3.16: The values of HE and $N_1, \ldots, N_m$ produced by the generalized state-space method when different numbers of q-forwards are used.

HE	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}
0.8425	83.1379	–	–	–	–	–	–	–	–	–
0.8517	9.6612	78.6630	–	–	–	–	–	–	–	–
0.8583	19.5117	4.3579	67.1848	–	–	–	–	–	–	–
0.8640	26.2820	4.5859	4.9870	57.6373	–	–	–	–	–	–
0.8685	30.9677	4.7438	5.2967	5.6566	49.0497	–	–	–	–	–
0.8717	34.1813	4.8520	5.5091	5.9726	6.2574	40.9674	–	–	–	–
0.8736	36.3330	4.9245	5.6513	6.1842	6.5404	6.7341	33.1843	–	–	–
0.8749	37.7136	4.9710	5.7425	6.3200	6.7220	6.9642	7.0601	25.6171	–	–
0.8754	38.5350	4.9987	5.7968	6.4008	6.8301	7.1011	7.2282	7.2236	18.2404	–
0.8758	38.9536	5.0128	5.8245	6.4420	6.8851	7.1709	7.3139	7.3269	7.2210	11.0539

In contrast to the q-forwards used in the baseline results, the q-forwards considered here are linked to distinct birth cohorts.

Table 3.17 shows Result I (a comparison of different hedging methods) on the basis of the alternative set of q-forwards. The conclusion obtained in Section 3.6.2 remains unchanged: the generalized state-space method consistently yields a better hedge effectiveness compared to the delta and delta-nuga methods, regardless of whether the original CBD model or the LLCBD model is assumed in the derivation.

Table 3.18 displays Result II (the impact of model mis-specification) on the basis of the alternative set of q-forwards. The conclusion drawn in Section 3.6.3 still stands: neglecting stochastic drifts may lead to a material loss of hedge effectiveness, if the true underlying model is one with stochastic drifts.

3.7 Concluding Remarks

Longevity risk comprises of diffusion risk and drift risk. Although both sources of risk are significant, the latter is often ignored in the existing stochastic mortality models. In this chapter, we introduce the LLCBD model which captures drift risk by allowing the drifts themselves to follow a random walk. Written in a state-space form, the LLCBD model contains four hidden states, $\kappa_1(t)$, $\kappa_2(t)$, $C_1(t)$ and $C_2(t)$, all of which have demographic intuitions behind. All hidden states and parameters in the model can be estimated in one single stage by using the EM algorithm and the Kalman filter.

As an illustration, we estimate the LLCBD model to the historical mortality data of Canadian males. The adequacy of the model's fit is confirmed by Harvey's (1990) test that

Table 3.17: Result I (a comparison of different hedging methods) based on the alternative set of q-forwards. For Groups 1 and 3, the simulation model used is the original CBD model. For Groups 2 and 4, the simulation model used is the LLCBD model.

Group 1 Assumed model: The original CBD model Method: Traditional delta hedging					Group 3 Assumed model: The original CBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
0.8510	152.5614	43.2934	–	–	0.9135	112.2606	35.4131	–	–
−1.6532	−378.3325	–	562.1118	–	0.7646	112.5175	–	137.0038	–
−1.2703	−168.3086	–	–	367.6104	0.7597	152.6432	–	–	97.0379
0.7050	–	30.8523	161.5324	–	0.9106	–	30.8712	70.1558	–
0.4764	–	22.7090	–	174.7847	0.9107	–	32.3435	–	51.5237
−2.8836	–	–	−450.4640	662.2052	0.7358	–	–	120.3346	48.6849
Group 2 Assumed model: The LLCBD model Method: Traditional delta/delta-nuga hedging					Group 4 Assumed model: The LLCBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
0.6880	178.7526	59.9580	–	–	0.7905	11.7794	57.6941	–	–
−9.1519	−469.6742	–	806.0370	–	0.7086	−238.5952	–	184.9745	–
−12.2549	−213.1537	–	–	575.5422	0.7694	−124.0271	–	–	118.4957
0.0151	–	43.4293	222.2012	–	0.8300	–	45.7134	51.2775	–
−1.6542	–	32.6105	–	262.5109	0.8929	–	37.2561	–	58.1600
−18.8856	–	–	−669.7701	1053.7844	0.7267	–	–	−44.6941	136.6734
0.5892	349.2159	66.6085	−399.0373	221.0896	0.9132	−20.9026	37.8663	−60.2340	96.9143

is based on the model's vector of prediction errors. In comparison to the original CBD model, the LLCBD model provides a better goodness-of-fit in terms of AIC, and yields more accurate short- and long-term forecasts in terms of ME and MSE. We also find that the LLCBD model generates forecasts that are more consistent with the observed trends in the recent past and are more robust relative to changes in the length of the calibration window. Because the LLCBD model incorporates additionally the uncertainty associated with the drifts, it results in wider long-term prediction intervals that reflect the possibilities of future trend changes.

Another contribution of this chapter is the generalized state-space hedging method, from which one can construct an index-based longevity hedge to mitigate both diffusion and drift risks. As explained in Section 3.5.5, the proposed hedging method can ameliorate the problems of sub-optimality and singularity that the traditional delta and delta-nuga hedging methods are subject to. The proposed hedging method does not impose any requirement on the number of hedging instruments used. It also works for any combination

Table 3.18: Result II (the impact of model mis-specification) derived using the alternative set of q-forwards. The simulation model used is the LLCBD model.

Group 3 Assumed model: The original CBD model Method: Generalized state-space hedging					Group 4 Assumed model: The LLCBD model Method: Generalized state-space hedging				
HE	N_1	N_2	N_3	N_4	HE	N_1	N_2	N_3	N_4
Number of q-forwards used: $m = 1$									
0.7413	–	42.9262	–	–	0.7899	–	57.7622	–	–
Number of q-forwards used: $m = 2$									
0.8825	–	32.3435	–	51.5237	0.8929	–	37.2561	–	58.1600
Number of q-forwards used: $m = 3$									
0.7906	86.7842	28.9276	–	39.8774	0.9123	–	39.0704	−72.4252	101.2808
Number of q-forwards used: $m = 4$									
0.7840	82.7462	28.4464	10.8472	34.4610	0.9132	−20.9026	37.8663	−60.2340	96.9143

of hedging instruments, provided that the payoffs from the hedging instruments are not perfectly correlated with one another.

As an example, we apply the generalized state-space hedging method to a hypothetical hedging scenario. The results of this application point to three conclusions. First, the proposed hedging method performs better than the traditional delta and delta-nuga hedging methods, no matter which of the two models under consideration is assumed. Second, ignoring stochastic drifts in the derivation of the hedging strategies would lead to a material reduction in hedge effectiveness, if the true underlying model is one with stochastic drifts. Third, the negative impact of ignoring stochastic drifts is particularly significant if the duration of the liability being hedged is long.

To focus on the issue concerning drift risk, in this chapter we choose to build our model on the simplest version of the CBD model, which does not take cohort effects into account. We acknowledge that cohort effects are significant in certain populations, and that it is not trivial to incorporate cohort effects in a state-space representation in which the vector of hidden states evolve over time rather than year of birth. In future research, it would be interesting to investigate how the LLCBD model can be further extended to incorporate cohort effects. This goal may possibly be accomplished by using the parsimonious approach introduced by Marvos et al. (2014), whereby cohort effects are captured through the dependence among residuals.

To ease exposition, we have also assumed that the q-forwards and the annuity liability are associated with exactly the same population. In reality, the q-forwards are more likely

to be linked to a broad-based mortality index rather than the hedger's own population, giving rise to population basis risk that may reduce hedge effectiveness. Future research warrants a study on how our contributions may be applied to situations when population basis risk exists. Such a study would encompass an extension of the LLCBD model to a version which, similar to the models contributed by Li and Lee (2005), Cairns et al. (2011b), Dowd et al. (2011) and Li et al. (2015b), models the mortality of two populations in a coherent manner. The study would also include an adjustment of the generalized state-space hedging method to capture the imperfect correlations between the mortality improvements of the two populations involved. The multi-population extension of the generalized state-space hedging method is discussed in Chapter 4.

The longevity hedging strategy presented in this chapter is static, as no adjustment is made to the hedge after inception. Static hedging strategies generally require longer-dated instruments, but the majority of capital market investors prefer to invest in shorter-dated ones. It would therefore be useful to extend the proposed hedging method from static to dynamic. This extension is also discussed in Chapter 4.

Chapter 4

A Hedging Method with an Explicit Measure of Population Basis Risk

4.1 Introduction

Longevity risk refers to the adverse financial consequences that arise when individuals live longer than expected. The risk affects sponsors of defined-benefit pension plans heavily, as the longer people live the larger the pension liabilities are. The threat of longevity risk to pension plan sponsors has become even more apparent in recent years, due in part to the low-yield environment after the financial crisis of 2007-08 and in part to the more conservative mortality improvement scales that are recently introduced by the actuarial profession (Canadian Institute of Actuaries, 2014; Continuous Mortality Investigation Bureau, 2009a,b; Society of Actuaries, 2014). In response to the problem, pension longevity risk transfers started to emerge in early 2000s, providing opportunities for pension plan sponsors to transfer longevity risk to entities that are in a better position to run the risk.

The majority of the pension longevity risk transfers executed to date are insurance-based buy-outs, buy-ins and bespoke longevity swaps. However, the insurance industry is unable to take an unlimited amount of longevity risk, because of the systematic nature of risk and the capital requirements imposed by Solvency II or its equivalent. The analyses of Graziani (2014) and Michaelson and Mulholland (2014) concluded that the global insurance industry can only absorb a fraction of the longevity risk exposure from defined-benefit pension plans worldwide. Consequently, there is a need to search for additional participants who are willing to accept the risk. One possible candidate is capital market investors, who are possibly interested in the longevity asset class because of the risk premium and diversification benefit it offers.

Capital market investors demand liquidity and are likely to be discouraged by the information asymmetry arising from the fact that pension plan sponsors know better about the mortality experience of the individuals associated with their portfolios. To attract their participation, longevity risk needs to be packaged as standardized index-based mortality derivatives. Several successful attempts of standardization have been made. The first took place in January 2008 when J.P. Morgan executed a mortality (q-) forward contract with Lucida, a monoline insurer in the UK.[1] Other examples include the £70 million 10-year q-forward that J.P. Morgan transacted with the Pall (UK) Pension fund in January 2011[2] and the €12 billion index-based longevity swap executed between Deutsche Bank and Aegon in February 2012.[3] More recently, Deutsche Bank launched in November 2013 the Longevity Experience Option (LEO), which is structured as an out-of-the-money call option spread on 10-year forward survival rates and has a maturity of 10 years. The first LEO was reportedly traded in January 2014.[4] At present, tradable longevity indexes include the LifeMetrics Index provided by the Life and Longevity Markets Association (LLMA) and the Xpect Club Vita Index provided by Deutsche Börse.

Nevertheless, the market for index-based mortality derivatives is still quite far from being large and liquid, and is facing several challenges. On the supply side, a few major investment banks, namely Credit Suisse, Normura and UBS, left the market for longevity risk transfers in 2012 for regulatory reasons (Tan et al., 2015). On the demand side, as noted by Blake et al. (2013), the major challenge is "the continuing resistance of pension plan trustees and their advisors, as well as insurers and reinsurers, to imperfect hedging solutions of the capital markets." Longevity hedges developed from index-based mortality derivatives are imperfect, because there exists population basis risk which arises from the difference in mortality improvements between the populations associated with the hedger and the index used for hedging purposes. The challenge on the demand side calls for research to address the following questions: (1) How to quantify the exposure to population basis risk in an index-based longevity hedge? (2) How to optimize an index-based longevity hedge when population basis risk is taken into account? (3) How risk management decisions, for example, selection of the most appropriate reference population, can be made in a world where population basis risk exists?

Although a large number of multi-population stochastic mortality models have been proposed,[5] only a few attempts have been made to investigate how such models can be used

[1]Source: www.efinancialnews.com/story/2008-02-19/lucida-guards-against-longevity

[2]Source: www.pensionsworld.co.uk/article/first-longevity-hedge-non-retired-pension-plan-members

[3]Source: www.db.com/medien/en/content/3862_4047.htm

[4]Source: www.trading-risk.com/deutsche-bank-longevity-option-platform-closes-debut-deal

[5]See, for example, D'Amato et al. (2014), Li and Lee (2005), Cairns et al. (2011b), Dowd et al. (2011b), Jarner and Kryger (2011), Ahmadi and Li (2014), Hatzopoulos and Haberman (2013), Hunt and Blake

to generate meaningful measures of population basis risk. Previous studies on measuring population basis risk, including the work of Cairns et al. (2014), Li and Hardy (2011) and Ngai and Sherris (2011), typically follow the framework set out by Coughlan et al. (2011). The framework encompasses the following two steps: (i) calibrate the portfolio of hedging instruments; (ii) calculate the amount of risk that the hedge can reduce on the basis of a large number of scenarios simulated from the assumed stochastic mortality model. The two steps are performed under the assumptions that population basis risk is absent and present, and the difference between the results under the two assumptions represents the population basis risk that the hedger is subject to. Because this approach depends on a calibrated longevity hedge, it does not readily indicate how population basis risk varies as the composition of the hedge portfolio changes. In addition, the heavy reliance on simulations makes this approach rather computationally demanding.

Strategies for optimizing an index-based longevity hedge have been developed by researchers including Cairns (2011, 2013), Cairns et al. (2006b, 2014), Coughlan et al. (2011), Dahl (2004), Dahl and Møller (2006), Dahl et al. (2008), Dowd et al. (2011a), Li and Hardy (2011), Li and Luo (2012), Luciano et al. (2012), and Zhou and Li (2014). Generally speaking, the hedging strategies were derived by matching the sensitivities of the liability being hedged and the portfolio of hedging instruments with respect to changes in the underlying mortality rates. Most of the existing longevity hedging strategies are subject to one significant limitation: they were developed under the ideal assumption that population basis risk does not exist. Although Dowd et al. (2011a), Li and Hardy (2011), Li and Luo (2012) and Zhou and Li (2014) studied how hedging strategies may be adjusted when population basis risk is present, their adjustment formulas are derived on the basis of specific multi-population mortality models. For example, the adjustment formula proposed by Li and Luo (2012) depends heavily on the augmented common factor model (Li and Lee, 2005) and is consequently incompatible with other model assumptions. A more general framework for hedging in the presence of population basis risk is yet to be sought.

The objectives of this chapter are threefold. The first objective is to develop a hedging strategy that explicitly incorporates population basis risk. This objective is achieved by extending the generalized state-space hedging method proposed in Chapter 3 to incorporate the random deviation between the mortality improvements of the populations associated with the hedging instruments and the liability being hedged. We choose to use the generalized state-space hedging method as a starting point, because, as explained in Chapter 3, the method ameliorates the problems of sub-optimality and singularity that are found in the delta-nuga hedging method of Cairns (2013) and imposes less stringent requirements on the number of hedging instruments used. In our proposed method, the hedging strategy

(2015), Yang and Wang (2013), and Zhou et al. (2013a, 2014).

is derived by first reformulating the assumed multi-population stochastic mortality model in a state-space form, and then considering the sensitivities of the hedge portfolio and the liability being hedged to all relevant hidden states. Our proposed method is rather general in the sense that it can be applied to any coherent multi-population mortality model, provided that the model can be expressed in a state-space form. We demonstrate in the next section that the applicable models include the multi-population versions of the Lee-Carter and Cairns-Blake-Dowd models (Lee and Carter, 1992; Cairns et al., 2006). The proposed hedging method is presented in both static and dynamic settings, and the benefit of dynamically adjusting a longevity hedge is studied.

The second objective is to develop an efficient method for quantifying the population basis risk in a longevity hedge. To this end, we introduce a method to analytically approximate the variance of the time-t values of an annuity portfolio, with or without a longevity hedge. The method enables us to estimate hedge effectiveness (defined as the reduction in variance) without using simulation. As such, we can readily gauge how much hedge effectiveness is eroded by population basis risk under different hedge portfolio compositions. We also show that the approximated variance of a hedged annuity portfolio can be decomposed into components arising solely from the hidden states that are shared by all populations and components stemming exclusively from the hidden states that are population-specific. The latter components collectively represent an explicit measure of the population basis risk involved in the longevity hedge. Furthermore, from the mathematical formulation of this measure, we can infer that a portion of population basis risk depends on how the longevity hedge is constructed while another portion exists no matter what the notional amounts of the hedging instruments are. The intuitions behind the decomposition of variance and the measure of population basis risk are illustrated extensively by graphical means throughout the chapter.

The final objective is to develop a metric for assessing the relative levels of population basis risk that different hedging instruments may lead to. We propose a metric called 'standardized basis risk profile', with which we can compare the resulting population basis risk when q-forwards with different reference ages, maturities and reference populations are used. The standardized basis risk profile is of practical importance. For instance, let us suppose that a pension plan sponsor in Canada is contemplating hedging its longevity risk exposure with q-forwards that are linked to a LifeMetrics Index. At present, the LifeMetrics Index is available in England and Wales, the US, Germany and the Netherlands, but not in Canada. The standardized basis risk profile can aid this pension plan sponsor in selecting one out of the four available reference populations. While the methods proposed previously by Cairns (2011, 2013), Cairns et al. (2014), Coughlan et al. (2011) and Li and Hardy (2011) may also be used for the purpose of reference population selection, they generally

require significant computational effort. In contrast, the standardized basis risk profile can be computed analytically, and the comparison between different standardized basis risk profiles can be made without calibrating the associated hedges in advance.

The remainder of this chapter is organized as follows. Section 4.2 describes the collection of multi-population mortality models to which the proposed hedging strategy can be applied. Section 4.3 presents the proposed hedging strategy and the analytical decomposition of the portfolio variance. Section 4.4 explains how the decomposition of variance can be utilized to measure population basis risk and defines the standardized basis risk profile. Section 4.5 illustrates the proposed methodologies using real mortality data from various national populations. Finally, Section 4.6 concludes the chapter.

4.2 The Applicable Multi-Population Mortality Models

The proposed hedging method is quite general in the sense that it can be applied to any coherent multi-population mortality model, provided that the model can be expressed in a state-space form. In what follows, we first present the general state-space representation, and then illustrate how existing coherent multi-population mortality models can be formulated in state-space forms.

4.2.1 The General State-Space Representation

Generally speaking, a state-space representation is composed of two components: (1) an observation equation that describes the connection between the observations and hidden states, and (2) a transition equation that captures the evolution of the hidden states through a Markov process. Let $\vec{y}_t$ and $\vec{\alpha}_t$ be the vectors of observations and hidden states at time t, respectively. The observation equation is defined as

$$\vec{y}_t = \vec{A} + B\vec{\alpha}_t + \vec{\epsilon}_t, \tag{4.1}$$

where $\vec{A}$ is a vector of constant terms, B is the design matrix which specifies the linear relationship between the observations $\vec{y}_t$ and the hidden states $\vec{\alpha}_t$, and $\vec{\epsilon}_t$ is a vector of observation errors that has a zero mean and a constant covariance matrix R. The transition equation is defined as

$$\vec{\alpha}_t = \vec{U} + D\vec{\alpha}_{t-1} + \vec{\eta}_t, \tag{4.2}$$

where $\vec{U}$ represents the vector of drifts, D is a matrix which specifies the relationship between $\vec{\alpha}_t$ and $\vec{\alpha}_{t-1}$, and $\vec{\eta}_t$ is a vector of innovations that has a zero mean and a constant covariance matrix Q.

In the context of multi-population mortality modeling, the observation vector $\vec{y}_t$ contains the time-t values of the (transformed) age-specific death rates or probabilities for all of the populations being modeled. Suppose that the number of populations being modeled is n_p. We can express $\vec{y}_t$ as $((\vec{y}_t^{(1)})', \ldots, (\vec{y}_t^{(n_p)})')'$, with $\vec{y}_t^{(p)}$, $p = 1, \ldots, n_p$, representing the vector of (transformed) age-specific death rates or probabilities for population p. If the age range under consideration is $[x_a, x_b]$, then $\vec{y}_t^{(p)} = (y_{x_a,t}^{(p)}, y_{x_a+1,t}^{(p)}, \ldots, y_{x_b,t}^{(p)})'$.

A typical multi-population mortality model contains several stochastic components, some of which apply to all of the populations being modeled while the other of which are population-specific. Accordingly, we can express the state vector as $\vec{\alpha}_t = ((\vec{\alpha}_t^c)', (\vec{\alpha}_t^{(1)})', \ldots, (\vec{\alpha}_t^{(n_p)})')'$, where $\vec{\alpha}_t^c$ encompasses the hidden states that are shared by all of the populations being modeled and $\vec{\alpha}_t^{(p)}$, $p = 1, \ldots, n_p$, consists of the hidden states that belong exclusively to population p. In the sequel, we call $\vec{\alpha}_t^c$ the vector of 'common states' and $\vec{\alpha}_t^{(p)}$ a vector of 'population-specific states'.

We require the multi-population mortality model to be 'coherent'. A coherent multi-population mortality model is constructed in such a way to avoid a long-term divergence between projected mortality trajectories of different populations, an outcome that is deemed to be biologically unreasonable. The concept of coherent multi-population mortality forecasting was first proposed by Li and Lee (2005), who also specified the following sufficient conditions for coherence: (i) all of the populations being modeled have the same age-response to the common stochastic factors; (ii) the population-specific stochastic factors are mean-reverting. The sufficient conditions for coherence guarantee that, the mortality trends for different populations would not diverge indefinitely.[6]

When formulated in a state-space representation, condition (i) means that the coefficients of the common states for all of the populations being modeled are the same.

[6]The concept of coherence has been translated into mathematical hypothesis by several researchers (e.g., Cairns et al., 2011b, Zhou and Li, 2014). Cairns et al. (2011b) modeled the log of central death rates, $\log(m_{x,t}^{(p)})$, using a two-population APC model. They applied the following mathematical hypothesis for coherence: the ratio of $m_{x,t}^{(1)}/m_{x,t}^{(2)}$ for all ages do not diverge as $t \to \infty$. In the M-CBD model proposed by Zhou and Li (2014), although it is not stated explicitly, the underlying mathematical hypothesis for coherence is that in logit scale, the death probabilities $q_{x,t}^{(1)}$ and $q_{x,t}^{(2)}$ do not diverge as $t \to \infty$.

Equivalently speaking, the design matrix B should satisfy the following structure:

$$B = \begin{pmatrix} B^c & B^{(1)} & 0 & \cdots & 0 \\ B^c & 0 & B^{(2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B^c & 0 & 0 & \cdots & B^{(n_p)} \end{pmatrix}, \tag{4.3}$$

where

$$B^c = \begin{pmatrix} (\vec{b}^c_{x_a})' \\ \vdots \\ (\vec{b}^c_{x_b})' \end{pmatrix} \quad \text{and} \quad B^{(p)} = \begin{pmatrix} (\vec{b}^{(p)}_{x_a})' \\ \vdots \\ (\vec{b}^{(p)}_{x_b})' \end{pmatrix}$$

for $p = 1, \ldots, n_p$. In the above, $\vec{b}^c_x$ is the vector of the coefficients of $\vec{\alpha}^c_t$ (the common states) in $y^{(p)}_{x,t}$ and does not depend on p; $\vec{b}^{(p)}_x$, $p = 1, \ldots, n_p$, is the vector of the coefficients of $\vec{\alpha}^{(p)}_t$ (the population-specific states) in $y^{(p)}_{x,t}$ and is allowed to depend on p. Condition (ii) means that $\vec{\alpha}^{(p)}_t$, $p = 1, \ldots, n_p$, follows a certain autoregressive process.

In line with the structure of the state vector, we have $\vec{\eta}_t = ((\vec{\eta}^c_t)', (\vec{\eta}^{(1)}_t)', \ldots, (\vec{\eta}^{(n_p)}_t)')'$, where $\vec{\eta}^c_t$ and $\vec{\eta}^{(p)}_t$, $p = 1, \ldots, n_p$, represent the innovation vectors in the processes for $\vec{\alpha}^c_t$ and $\vec{\alpha}^{(p)}_t$, respectively. It is assumed that $\vec{\eta}^c_t$, $\vec{\eta}^{(1)}_t$, ..., $\vec{\eta}^{(n_p)}_t$ are mutually independent. Given this assumption, the covariance matrix Q in the transition equation is a block diagonal matrix,

$$Q = \begin{pmatrix} Q^c & 0 & \cdots & 0 \\ 0 & Q^{(1)} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & Q^{(n_p)} \end{pmatrix}, \tag{4.4}$$

with $Q^c, Q^{(1)}, \ldots, Q^{(n_p)}$ representing the covariance matrices of $\vec{\eta}^c_t, \vec{\eta}^{(1)}_t, \ldots, \vec{\eta}^{(n_p)}_t$, respectively. In the rest of this section, we use two examples to illustrate how existing coherent multi-population mortality models can be formulated using the general state-space representation.

4.2.2 The Augmented Common Factor Model

The first example is the Augmented Common Factor (ACF) model proposed by Li and Lee (2005). Let $m_{x,t}^{(p)}$ be population p's central death rate at age x and in year t. The ACF model assumes that

$$\ln(m_{x,t}^{(p)}) = a_x^{(p)} + b_x^c \, k_t^c + b_x^{(p)} \, k_t^{(p)} + \epsilon_{x,t}^{(p)}, \qquad p = 1, \ldots, n_p,$$

where $a_x^{(p)}$ is an age-specific parameter representing population p's average mortality level at age x, k_t^c is a time-varying stochastic factor that affects all of the n_p populations, $k_t^{(p)}$ is another time-varying stochastic factor that affects population p only, and b_x^c and $b_x^{(p)}$ measure the sensitivities of $\ln(m_{x,t}^{(p)})$ to k_t^c and $k_t^{(p)}$, respectively. It is assumed that the error term $\epsilon_{x,t}^{(p)}$ is normally distributed with a zero mean and a variance of σ_ϵ^2, and that the error terms for different populations, ages and years are independent.

The trend in k_t^c can be interpreted to mean the general trend in mortality improvements for all of the n_p populations. As in the original Lee-Carter model (Lee and Carter, 1992), the evolution of k_t^c over time is modeled by a random walk with drift:

$$k_t^c = \mu^c + k_{t-1}^c + \eta_t^c,$$

where μ^c is the drift term and $\eta_t^c \overset{\text{i.i.d.}}{\sim} N(0, Q^c)$.

Deviations from the general trend are captured by the dynamics of $k_t^{(p)}$, $p = 1, \ldots, n_p$. It is assumed that the evolution of $k_t^{(p)}$ follows a first order autoregressive process:

$$k_t^{(p)} = \mu^{(p)} + \phi^{(p)} k_{t-1}^{(p)} + \eta_t^{(p)},$$

where $\mu^{(p)}$ is a constant, $\phi^{(p)}$ is another constant with an absolute value that is strictly less than 1, and $\eta_t^{(p)} \overset{\text{i.i.d.}}{\sim} N(0, Q^{(p)})$. It is further assumed that $\eta_t^c, \eta_t^{(1)}, \ldots, \eta_t^{(n_p)}$ are mutually uncorrelated. The assumed process for $k_t^{(p)}$ ensures the resulting mortality projections are coherent.

Suppose that the ACF model is estimated to data over the age range of $[x_a, x_b]$. We can easily express such a model in a state-space form, with the vector of observations being

$$\vec{y}_t = ((\vec{y}_t^{(1)})', \ldots, (\vec{y}_t^{(n_p)})')' = \left(\ln(m_{x_a,t}^{(1)}), \ldots, \ln(m_{x_b,t}^{(1)}), \ldots, \ln(m_{x_a,t}^{(n_p)}), \ldots, \ln(m_{x_b,t}^{(n_p)})\right)',$$

where $\vec{y}_t^{(p)} = (\ln(m_{x_a,t}^{(p)}), \ldots, \ln(m_{x_b,t}^{(p)}))'$ for $p = 1, \ldots, n_p$, and the vector of hidden states

being

$$\vec{\alpha}_t = ((\vec{\alpha}_t^c)', (\vec{\alpha}_t^{(1)})', \ldots, (\vec{\alpha}_t^{(n_p)})')' = \left(k_t^c,\, k_t^{(1)},\, \ldots,\, k_t^{(n_p)}\right)',$$

where $\vec{\alpha}_t^c = k_t^c$ and $\vec{\alpha}_t^{(p)} = k_t^{(p)}$ for $p = 1, \ldots, n_p$.

The rest of the observation equation is specified as

$$\vec{A} = \left(a_{x_a}^{(1)}, \ldots, a_{x_b}^{(1)}, \ldots, a_{x_a}^{(n_p)}, \ldots, a_{x_b}^{(n_p)}\right)',$$

$$\vec{\epsilon}_t = \left(\epsilon_{x_a,t}^{(1)},\, \ldots,\, \epsilon_{x_b,t}^{(1)},\, \ldots, \epsilon_{x_a,t}^{(n_p)},\, \ldots,\, \epsilon_{x_b,t}^{(n_p)}\right)',$$

$$B^c = ((\vec{b}_{x_a}^c)', \ldots, (\vec{b}_{x_b}^c)')' = (b_{x_a}^c, \ldots, b_{x_b}^c)',$$

and

$$B^{(p)} = ((\vec{b}_{x_a}^{(p)})', \ldots, (\vec{b}_{x_b}^{(p)})')' = (b_{x_a}^{(p)}, \ldots, b_{x_b}^{(p)})',$$

where $\vec{b}_x^c = b_x^c$ and $\vec{b}_x^{(p)} = b_x^{(p)}$ for $p = 1, \ldots, n_p$. Note that $\vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, R)$, where $R = \sigma_\epsilon^2 \cdot \text{I}_{n_p(x_b - x_a + 1)}$ and $\text{I}_{n_p(x_b - x_a + 1)}$ represents an $n_p(x_b - x_a + 1) \times n_p(x_b - x_a + 1)$ identity matrix.

Finally, the rest of the transition equation is specified as

$$\vec{U} = \left(\mu^c, \mu^{(1)}, \ldots, \mu^{(n_p)}\right)',$$

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \phi^{(1)} & 0 & \cdots & 0 \\ 0 & 0 & \phi^{(2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \phi^{(n_p)} \end{pmatrix},$$

and

$$\vec{\eta}_t = ((\vec{\eta}_t^c)', (\vec{\eta}_t^{(1)})', \ldots, (\vec{\eta}_t^{(n_p)})')' = (\eta_t^c, \eta_t^{(1)}, \ldots, \eta_t^{(n_p)})',$$

where $\vec{\eta}_t^c = \eta_t^c$ and $\vec{\eta}_t^{(p)} = \eta_t^{(p)}$ for $p = 1, \ldots, n_p$. Note that $\vec{\eta}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q)$, where Q is specified by equation (4.4) with Q^c being the variance of η_t^c and $Q^{(p)}$ being the variance of $\eta_t^{(p)}$.

4.2.3 The Multi-Population Cairns-Blake-Dowd Model

The second example is the multi-population Cairns-Blake-Dowd (M-CBD) model, which Zhou and Li (2014) considered. Let $q_{x,t}^{(p)}$ represents the probability that an individual from population p dies between time $t-1$ and t (during year t), provided that he or she has survived to age x at time $t-1$. The M-CBD model assumes that

$$\text{logit}(q_{x,t}^{(p)}) = \ln\left(\frac{q_{x,t}^{(p)}}{1-q_{x,t}^{(p)}}\right) = \kappa_{1,t}^{c} + \kappa_{2,t}^{c}(x-\bar{x}) + \kappa_{1,t}^{(p)} + \kappa_{2,t}^{(p)}(x-\bar{x}) + \epsilon_{x,t}^{(p)}, \qquad p = 1, \ldots, n_p,$$

where $\bar{x}$ denotes the average age over the sample age range $[x_a, x_b]$, $\kappa_{1,t}^{c}$ and $\kappa_{2,t}^{c}$ are time-varying stochastic factors that apply to all of the n_p populations, $\kappa_{1,t}^{(p)}$ and $\kappa_{2,t}^{(p)}$ are time-varying stochastic factors that apply to population p only, and $\epsilon_{x,t}^{(p)}$ is the error term. As in the ACF model, it is assumed that $\epsilon_{x,t}^{(p)} \sim N(0, \sigma_\epsilon^2)$ and that the error terms for different populations, ages and years are independent.

The common stochastic factors $\kappa_{1,t}^{c}$ and $\kappa_{2,t}^{c}$ are modeled jointly by a bivariate random walk with drifts:

$$\begin{cases} \kappa_{1,t}^{c} &= \mu_1^c + \kappa_{1,t-1}^{c} + \eta_{1,t}^{c} \\ \kappa_{2,t}^{c} &= \mu_2^c + \kappa_{2,t-1}^{c} + \eta_{2,t}^{c} \end{cases},$$

where μ_1^c and μ_2^c are the drift terms and $\vec{\eta}_t^{\,c} = (\eta_{1,t}^{c}, \eta_{2,t}^{c})' \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q^c)$.

Each pair of population-specific stochastic factors, $\kappa_{1,t}^{(p)}$ and $\kappa_{2,t}^{(p)}$, are modeled by two correlated first-order autoregressive processes:

$$\begin{cases} \kappa_{1,t}^{(p)} &= \mu_1^{(p)} + \phi_1^{(p)}\kappa_{1,t-1}^{(p)} + \eta_{1,t}^{(p)} \\ \kappa_{2,t}^{(p)} &= \mu_2^{(p)} + \phi_2^{(p)}\kappa_{2,t-1}^{(p)} + \eta_{2,t}^{(p)} \end{cases},$$

where $\mu_1^{(p)}$ and $\mu_2^{(p)}$ are constants, $\phi_1^{(p)}$ and $\phi_2^{(p)}$ are constants with absolute values that are strictly less than 1, and $\vec{\eta}_t^{\,(p)} = (\eta_{1,t}^{(p)}, \eta_{2,t}^{(p)})' \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q^{(p)})$. The use of mean-reverting processes for $\kappa_{1,t}^{(p)}$ and $\kappa_{2,t}^{(p)}$ ensures that the resulting mortality forecasts are coherent.

The M-CBD model can also be written easily in a state-space form, with the vector of observations being

$$\vec{y}_t = ((\vec{y}_t^{\,(1)})', \ldots, (\vec{y}_t^{\,(n_p)})')' = \left(\text{logit}(q_{x_a,t}^{(1)}), \ldots, \text{logit}(q_{x_b,t}^{(1)}), \ldots, \text{logit}(q_{x_a,t}^{(n_p)}), \ldots, \text{logit}(q_{x_b,t}^{(n_p)})\right)',$$

where $\vec{y}_t^{(p)} = (\text{logit}(q_{x_a,t}^{(p)}), \ldots, \text{logit}(q_{x_b,t}^{(p)}))'$, and the vector of hidden states being

$$\vec{\alpha}_t = ((\vec{\alpha}_t^c)', (\vec{\alpha}_t^{(1)})', \ldots, (\vec{\alpha}_t^{(n_p)})')' = \left(\kappa_{1,t}^c, \kappa_{2,t}^c, \kappa_{1,t}^{(1)}, \kappa_{2,t}^{(1)}, \ldots, \kappa_{1,t}^{(n_p)}, \kappa_{2,t}^{(n_p)}\right)',$$

where $\vec{\alpha}_t^c = (\kappa_{1,t}^c, \kappa_{2,t}^c)'$ and $\vec{\alpha}_t^{(p)} = (\kappa_{1,t}^{(p)}, \kappa_{2,t}^{(p)})'$ for $p = 1, \ldots, n_p$.

The rest of the observation equation is specified as

$$\vec{A} = (0, \ldots, 0)',$$

$$\vec{\epsilon}_t = \left(\epsilon_{x_a,t}^{(1)}, \ldots, \epsilon_{x_b,t}^{(1)}, \ldots, \epsilon_{x_a,t}^{(n_p)}, \ldots, \epsilon_{x_b,t}^{(n_p)}\right)',$$

and

$$B^c = \begin{pmatrix} (\vec{b}_{x_a}^c)' \\ (\vec{b}_{x_a+1}^c)' \\ \vdots \\ (\vec{b}_{x_b}^c)' \end{pmatrix} = B^{(p)} = \begin{pmatrix} (\vec{b}_{x_a}^{(p)})' \\ (\vec{b}_{x_a+1}^{(p)})' \\ \vdots \\ (\vec{b}_{x_b}^{(p)})' \end{pmatrix} = \begin{pmatrix} 1 & x_a - \bar{x} \\ 1 & x_a + 1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_b - \bar{x} \end{pmatrix}, \quad \text{for} \quad p = 1, 2, \ldots, n_p,$$

where $\vec{b}_x^c = \vec{b}_x^{(p)} = (1, x - \bar{x})'$, for $p = 1, \ldots, n_p$ and $x = x_a, \ldots, x_b$. Note that $\vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, R)$, where $R = \sigma_\epsilon^2 \cdot \text{I}_{n_p(x_b - x_a + 1)}$.

Finally, the rest of the transition equation is specified as

$$\vec{U} = \left(\mu_1^c, \mu_2^c, \mu_1^{(1)}, \mu_2^{(1)} \ldots, \mu_1^{(n_p)}, \mu_2^{(n_p)}\right)',$$

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \phi_1^{(1)} & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \phi_2^{(1)} & \cdots & 0 & 0 \\ \vdots & \vdots & & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \phi_1^{(n_p)} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \phi_2^{(n_p)} \end{pmatrix}$$

and

$$\vec{\eta}_t = ((\vec{\eta}_t^c)', (\vec{\eta}_t^{(1)})', \ldots, (\vec{\eta}_t^{(n_p)})')' = (\eta_{1,t}^c, \eta_{2,t}^c, \ldots, \eta_{1,t}^{(n_p)}, \eta_{2,t}^{(n_p)})'.$$

Note that $\vec{\eta}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q)$, where Q is specified by equation (4.4) with Q^c being the covariance matrix of $\vec{\eta}_t^c$ and $Q^{(p)}$ being the covariance matrix of $\vec{\eta}_t^{(p)}$.

4.3 The Generalized State-Space Hedging Method

4.3.1 The Set-up

Let us first define several notations. Denote by $\mathcal{F}_t$ the information concerning the evolution of mortality up to and including time t. We use $m_{x,s|t}^{(p)}$ and $q_{x,s|t}^{(p)}$ to represent $m_{x,s}^{(p)}$ and $q_{x,s}^{(p)}$ given $\mathcal{F}_t$, respectively. If $s > t$, then $m_{x,s|t}^{(p)}$ and $q_{x,s|t}^{(p)}$ are random variables conditioned on the realized mortality up to and including time t. Otherwise, $m_{x,s|t}^{(p)}$ and $q_{x,s|t}^{(p)}$ are known realizations. Similarly, we use $\vec{\alpha}_{s|t}$, $\vec{\alpha}_{s|t}^c$, $\vec{\alpha}_{s|t}^{(p)}$ and $y_{x,s|t}^{(p)}$ to denote $\vec{\alpha}_s$, $\vec{\alpha}_s^c$, $\vec{\alpha}_s^{(p)}$ and $y_{x,s}^{(p)}$ given $\mathcal{F}_t$, respectively.

Suppose that it is now time t_0 (i.e., the end of year t_0). The hedger wishes to hedge the longevity risk associated with a portfolio of T-year temporary life annuity immediate contracts that are just sold to individuals aged x_0. We use P_L to denote the hedger's own population of individuals, and let $L(t)$ be the time-t value of the hedger's future liabilities (per annuitant), given the information about the evolution of mortality up to and including time t. Ignoring small sample risk, the value of $L(t)$ can be expressed as

$$\begin{aligned} L(t) &= \sum_{u=1}^{T-(t-t_0)} e^{-ru} \left(\prod_{s=t+1}^{t+u} (1 - q_{x_0+s-t_0-1,s|t}^{(P_L)}) \right) \\ &= \sum_{u=1}^{T-(t-t_0)} e^{-ru} \left(\prod_{s=t+1}^{t+u} g_{x_0+s-t_0-1,s|t}^{(P_L)} \right), \end{aligned}$$

for $t = t_0, t_0+1, \ldots, t_0+T-1$, where $g_{x,s|t}^{(p)}$ is a function that links $(1 - q_{x,s|t}^{(p)})$ to the observation being modeled. For the M-CBD model, $g_{x,s|t}^{(p)}$ is simply

$$g_{x,s|t}^{(p)} = 1 - q_{x,s|t}^{(p)} = \frac{1}{1 + \exp(y_{x,s|t}^{(p)})}.$$

For the ACF model, which is built for the central death rates, an assumption is needed to connect $m_{x,s|t}^{(p)}$ and $q_{x,s|t}^{(p)}$. We assume that

$$m_{x,s|t}^{(p)} = -\ln(1 - q_{x,s|t}^{(p)}),$$

which holds exact if the force of mortality is constant between consecutive integer ages. Under this assumption, $g^{(p)}_{x,s|t}$ for the ACF model is given by

$$g^{(p)}_{x,s|t} = 1 - q^{(p)}_{x,s|t} = \exp(-m^{(p)}_{x,s|t}) = \exp(-\exp(y^{(p)}_{x,s|t})).$$

When $t = t_0$, $L(t)$ represents the total value of the future liabilities payable to all annuitants. For $t_0 < t < t_0 + T$, $L(t)$ represents the value of the future liabilities payable to the annuitants who survive to time t. At time $t_0 + T$, the liabilities run off completely and hence $L(t_0 + T) = 0$.

The hedging instruments used are q-forwards. A q-forward is a zero-coupon swap with its floating leg proportional to the realized death probability at a certain reference age during the year immediately before the maturity date and its fixed leg proportional to the corresponding pre-determined forward mortality rate. The hedger should participate in the q-forwards as the fixed-rate receiver, so that if mortality turns out to be lower than expected, then it will receive a net payment to offset the correspondingly larger annuity liability. In general, the reference populations to which the q-forwards available in the market are linked are not identical to the hedger's own population of individuals. The hedger is therefore subject to population basis risk, which is addressed in later parts of this chapter.

We consider a hedge portfolio of $m \geq 1$ q-forwards, which are linked to the same reference population P_H. Let us focus on the jth q-forward contract, which has a reference age x_j and time-to-maturity T_j. If the q-forward is launched at time t, then its floating leg would be proportional to the realization of $q^{(P_H)}_{x_j,t+T_j|t}$, whereas its fixed leg would be proportional to the forward mortality rate $\mathrm{E}(q^{(P_H)}_{x_j,t+T_j|t})$ that is fixed at time t.[7] It follows that the (random) time-t value of its payoff to the fixed-rate receiver is

$$\begin{aligned} H^{(P_H)}_j(t) &= e^{-rT_j}\left(\mathrm{E}(q^{(P_H)}_{x_j,t+T_j|t}) - q^{(P_H)}_{x_j,t+T_j|t}\right) \\ &= e^{-rT_j}\left(\mathrm{E}(q^{(P_H)}_{x_j,t+T_j|t}) - 1 + g^{(P_H)}_{x_j,t+T_j|t}\right) \end{aligned}$$

per \$1 notional.

We use $N^{(P_H)}_j(t)$ to denote the notional amount of the jth q-forward acquired at time t. Our hedging goal is to minimize the variance of the time-t value of the unexpected

[7] In practice, the forward mortality rate should be lower than $\mathrm{E}(q^{(P_H)}_{x_j,t+T_j|t})$, so that the counterparty accepting longevity risk would be rewarded a positive expected risk premium. However, because the cost of hedging is not a focus of this chapter, we assume a zero risk premium for simplicity.

future cash flows. If the hedge is static, that is, the hedge portfolio is formed at $t = t_0$ and remains unadjusted thereafter, then the hedging goal can be formulated mathematically as

$$\min_{N_1^{(P_H)}(t_0),\ldots,N_m^{(P_H)}(t_0)} \left(\text{Var} \left(L(t_0) - \sum_{j=1}^{m} N_j^{(P_H)}(t_0) H_j^{(P_H)}(t_0) \right) \right). \tag{4.5}$$

We also consider dynamic hedging. In formulating a dynamic hedge, we assume that the hedger adjusts its hedge portfolio at the end of each year. We also assume that when the hedge is adjusted, the existing q-forwards are unwinded and freshly launched q-forwards (also with reference ages $x_1, \ldots x_m$ and times-to-maturity $T_1, \ldots T_m$) are introduced to the hedge portfolio. Under these assumptions, the time-t hedging goal for the dynamic hedge can be expressed as

$$\min_{N_1^{(P_H)}(t),\ldots,N_m^{(P_H)}(t)} \left(\text{Var} \left(L(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t) \right) \right), \tag{4.6}$$

for $t = t_0, t_0 + 1, \ldots, t_0 + T - 1$.

4.3.2 Decomposition of Variance

Before we derive the optimal notional amounts, we demonstrate in this sub-section that the variances in equations (4.5) and (4.6) can be decomposed into different components, each of which carries a specific physical meaning. The decomposition allows us to better understand the composition of the longevity risk that the hedger is subject to.

Given the state-space representation of the underlying multi-population mortality model, both $L(t)$ and $H_j^{(P_H)}(t)$ are functions of $\{\vec{\alpha}_s, s > t|\mathcal{F}_t\}$. Since $L(t)$ represents the value of the liabilities after time t until they run off completely at time $t_0 + T$, it contains the common states and the P_L-specific states at times $t+1, \ldots, t_0+T$; accordingly, $L(t)$ is a function of $\vec{\alpha}^c_{t+1|t}, \vec{\alpha}^c_{t+2|t}, \ldots, \vec{\alpha}^c_{t_0+T|t}$ and $\vec{\alpha}^{(P_L)}_{t+1|t}, \vec{\alpha}^{(P_L)}_{t+2|t}, \ldots, \vec{\alpha}^{(P_L)}_{t_0+T|t}$. On the other hand, because $H_j^{(P_H)}(t)$ represents the value of the payoff from the jth q-forward at time $t + T_j$, $H_j^{(P_H)}(t)$ contains the common state and the P_H-specific state at time $t + T_j$; hence, $H_j^{(P_H)}(t)$ is a function of $\vec{\alpha}^c_{t+T_j|t}$ and $\vec{\alpha}^{(P_H)}_{t+T_j|t}$.

To facilitate analytical calculations, we approximate $L(t)$ and $H_j^{(P_H)}(t)$ using first order Taylor expansions about the relevant states vectors. For $L(t)$, the first-order approximation

$l(t)$ is given by

$$\begin{aligned} L(t) &\approx l(t) \\ &= \hat{L}(t) + \sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' (\vec{\alpha}^c_{s|t} - \hat{\vec{\alpha}}^c_{s|t}) + \sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{s|t}}\right)' (\vec{\alpha}^{(P_L)}_{s|t} - \hat{\vec{\alpha}}^{(P_L)}_{s|t}), \end{aligned}$$

where $\hat{\vec{\alpha}}^c_{s|t}$ and $\hat{\vec{\alpha}}^{(P_L)}_{s|t}$ are the expected values of $\vec{\alpha}^c_s$ and $\vec{\alpha}^{(P_L)}_s$ given $\mathcal{F}_t$, respectively, and $\hat{L}(t)$ is the value of $L(t)$ evaluated at $\hat{\vec{\alpha}}^c_{t+1|t}, \ldots, \hat{\vec{\alpha}}^c_{t_0+T|t}$ and $\hat{\vec{\alpha}}^{(P_L)}_{t+1|t}, \ldots, \hat{\vec{\alpha}}^{(P_L)}_{t_0+T|t}$. For $H^{(P_H)}_j(t)$, $j = 1, \ldots, m$, the first order approximation $h^{(P_H)}_j(t)$ is given by

$$\begin{aligned} H^{(P_H)}_j(t) &\approx h^{(P_H)}_j(t) \\ &= \hat{H}^{(P_H)}_j(t) + \left(\frac{\partial H^{(P_H)}_j(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right)' (\vec{\alpha}^c_{t+T_j|t} - \hat{\vec{\alpha}}^c_{t+T_j|t}) + \left(\frac{\partial H^{(P_H)}_j(t)}{\partial \vec{\alpha}^{(P_H)}_{t+T_j|t}}\right)' (\vec{\alpha}^{(P_H)}_{t+T_j|t} - \hat{\vec{\alpha}}^{(P_H)}_{t+T_j|t}), \end{aligned}$$

where $\hat{\vec{\alpha}}^c_{t+T_j|t}$ and $\hat{\vec{\alpha}}^{(P_H)}_{t+T_j|t}$ are the expected values of $\vec{\alpha}^c_{t+T_j|t}$ and $\vec{\alpha}^{(P_H)}_{t+T_j|t}$ given $\mathcal{F}_t$, respectively, and $\hat{H}^{(P_H)}_j(t)$ is the value of $H^{(P_H)}_j(t)$ evaluated at $\hat{\vec{\alpha}}^c_{t+T_j|t}$ and $\hat{\vec{\alpha}}^{(P_H)}_{t+T_j|t}$.

As with $L(t)$ and $H^{(P_H)}_j(t)$, $l(t)$ is a function of $\vec{\alpha}^c_{t+1|t}, \ldots, \vec{\alpha}^c_{t_0+T|t}$ and $\vec{\alpha}^{(P_L)}_{t+1|t}, \ldots, \vec{\alpha}^{(P_L)}_{t_0+T|t}$, and $h^{(P_H)}_j(t)$ is a function of $\vec{\alpha}^c_{t+T_j|t}$ and $\vec{\alpha}^{(P_H)}_{t+T_j|t}$. For brevity, these arguments are suppressed in the notations. Further, the partial derivatives $\partial L(t)/\partial \vec{\alpha}^c_{s|t}$, $\partial L(t)/\partial \vec{\alpha}^{(P_L)}_{s|t}$, $\partial H^{(P_H)}_j(t)/\partial \vec{\alpha}^c_{s|t}$ and $\partial H^{(P_H)}_j(t)/\partial \vec{\alpha}^{(P_H)}_{s|t}$ are evaluated at $\vec{\alpha}_{s|t} = \hat{\vec{\alpha}}_{s|t} = (\hat{\vec{\alpha}}^c_{s|t}, \hat{\vec{\alpha}}^{(1)}_{s|t}, \ldots, \hat{\vec{\alpha}}^{(n_p)}_{s|t})'$, unless otherwise specified.

We let $\vec{N}^{(P_H)}_t = (N^{(P_H)}_1(t), \ldots, N^{(P_H)}_m(t))'$ be the vector of notional amounts. Without loss of generality, we assume that $T_1 < T_2 < \ldots < T_m$. The variance to be minimized can be expressed as

$$\begin{aligned} &\mathrm{Var}\left(L(t) - \sum_{j=1}^{m} N^{(P_H)}_j(t) H^{(P_H)}_j(t)\right) \\ \approx &\,\mathrm{Var}\left(l(t) - \sum_{j=1}^{m} N^{(P_H)}_j(t) h^{(P_H)}_j(t)\right) \\ = &\,\mathrm{Var}\left(\sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' (\vec{\alpha}^c_{s|t} - \hat{\vec{\alpha}}^c_{s|t}) - \sum_{j=1}^{m} N^{(P_H)}_j(t) \left(\frac{\partial H^{(P_H)}_j(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right)' (\vec{\alpha}^c_{t+T_j|t} - \hat{\vec{\alpha}}^c_{t+T_j|t})\right. \\ &\left. + \sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{s|t}}\right)' (\vec{\alpha}^{(P_L)}_{s|t} - \hat{\vec{\alpha}}^{(P_L)}_{s|t}) - \sum_{j=1}^{m} N^{(P_H)}_j(t) \left(\frac{\partial H^{(P_H)}_j(t)}{\partial \vec{\alpha}^{(P_H)}_{t+T_j|t}}\right)' (\vec{\alpha}^{(P_H)}_{t+T_j|t} - \hat{\vec{\alpha}}^{(P_H)}_{t+T_j|t})\right). \end{aligned}$$

Noting that the common states, the states specific to the hedger's population and the states specific to the q-forward's reference population are mutually independent, we can express the variance above as the sum of five components:

$$\text{Var}\left(l(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t) h_j^{(P_H)}(t)\right) = V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t), \tag{4.7}$$

where

$$V_1(t) = \text{Var}\left(\sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' (\vec{\alpha}^c_{s|t} - \hat{\vec{\alpha}}^c_{s|t})\right) = \sum_{s,u=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' \text{Cov}(\vec{\alpha}^c_{s|t}, \vec{\alpha}^c_{u|t}) \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{u|t}}\right)$$

represents the portion of the variance of $L(t)$ (i.e., the time-t value of the hedger's future liabilities) that is contributed from the common states,

$$\begin{aligned} V_2(t) &= \text{Var}\left(\sum_{j=1}^{m} \left(-N_j^{(P_H)}(t) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right)' (\vec{\alpha}^c_{t+T_j|t} - \hat{\vec{\alpha}}^c_{t+T_j|t})\right) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} \left(-N_i^{(P_H)}(t) \frac{\partial H_i^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_i|t}}\right)' \text{Cov}(\vec{\alpha}^c_{t+T_i|t}, \vec{\alpha}^c_{t+T_j|t}) \left(-N_j^{(P_H)}(t) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right) \end{aligned}$$

represents the portion of the variance of $\sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ (i.e., the time-t value of the hedge portfolio) that is contributed from the common states,

$$\begin{aligned} V_3(t) &= 2\text{Cov}\left(\sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' (\vec{\alpha}^c_{s|t} - \hat{\vec{\alpha}}^c_{s|t}), \sum_{j=1}^{m} \left(-N_j^{(P_H)}(t) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right)' (\vec{\alpha}^c_{t+T_j|t} - \hat{\vec{\alpha}}^c_{t+T_j|t})\right) \\ &= 2\sum_{s=t+1}^{t_0+T}\sum_{j=1}^{m} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}}\right)' \text{Cov}(\vec{\alpha}^c_{s|t}, \vec{\alpha}^c_{t+T_j|t}) \left(-N_j^{(P_H)}(t) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_j|t}}\right) \end{aligned}$$

represents (the negative of) the covariance between $\sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ and $L(t)$,[8]

$$V_4(t) = \text{Var}\left(\sum_{s=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{s|t}}\right)' (\vec{\alpha}^{(P_L)}_{s|t} - \hat{\vec{\alpha}}^{(P_L)}_{s|t})\right) = \sum_{s,u=t+1}^{t_0+T} \left(\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{s|t}}\right)' \text{Cov}(\vec{\alpha}^{(P_L)}_{s|t}, \vec{\alpha}^{(P_L)}_{u|t}) \left(\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{u|t}}\right)$$

represents the portion of the variance of $L(t)$ that is contributed from the states that are

[8]Because the states specific to P_H and P_L are independent, the covariance between $\sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ and $L(t)$ is contributed entirely by the common states.

specific to the hedger's population,

$$\begin{aligned} V_5(t) &= \mathrm{Var}\left(-\sum_{j=1}^{m} N_j^{(P_H)}(t) \left(\frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^{(P_H)}} \right)' (\vec{\alpha}_{t+T_j|t}^{(P_H)} - \vec{\hat{\alpha}}_{t+T_j|t}^{(P_H)}) \right) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \left(-N_i^{(P_H)}(t) \frac{\partial H_i^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_i|t}^{(P_H)}} \right)' \mathrm{Cov}(\vec{\alpha}_{t+T_i|t}^{(P_H)}, \vec{\alpha}_{t+T_j|t}^{(P_H)}) \left(-N_j^{(P_H)}(t) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^{(P_H)}} \right) \end{aligned}$$

represents the portion of the variance of $\sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ that is contributed from the states that are specific to the q-forwards' reference population. It is noteworthy that $V_1(t), V_2(t), V_4(t)$ and $V_5(t)$ are all non-negative, because they can be regarded as the variances of certain random variables. However, $V_3(t)$ is negative (equivalently speaking, the covariance between $\sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ and $L(t)$ is positive), because the changes in the values of the q-forward portfolio and the liability offset each other.

We now drill deeper into the physical meanings behind the five variance components. First, let us focus on $V_1(t)$, $V_2(t)$ and $V_3(t)$, which are related to the common states but are independent of the population-specific states. The sum of $V_1(t)$, $V_2(t)$ and $V_3(t)$ can thus be interpreted as the portion of the variance of $L(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ (i.e., the time-t value of the hedged position) contributed from the common states. This portion of variance exists even if population basis risk is (hypothetically) absent.

In more detail, $V_1(t)$ is a constant that does not depend on the q-forwards' notional amounts, $V_2(t)$ is positively related to the q-forwards' notional amounts in a quadratic manner, and $V_3(t)$ is negatively related to the q-forwards' notional amounts in a linear manner. The pattern of $V_2(t)$ reflects the fact that the hedger is exposed to more common trend risk (i.e. the risk associated with the common states) as it acquires more notional amounts of q-forwards, while the pattern of $V_3(t)$ reflects the reduction in risk due to the correlation between the value of the liabilities being hedged and the value of the hedging instruments. If the liabilities are left unhedged, then both $V_2(t)$ and $V_3(t)$ would be zero. As the hedger acquires q-forwards, $V_3(t)$ decreases (becomes more negative) but $V_2(t)$ increases. If the decrease in $V_3(t)$ outweighs the increase in $V_2(t)$, then the portion of the variance that is contributed from the common states would be reduced (to a level that is smaller than $V_1(t)$).

Next, we turn to $V_4(t)$ and $V_5(t)$. Because these two components are related to the population-specific states but are independent of the common states, they collectively measure the population basis risk that the hedger is subject to. More specifically, as $V_4(t)$ is related only to $\vec{\alpha}_{t+1|t}^{(P_L)}, \ldots, \vec{\alpha}_{t_0+T|t}^{(P_L)}$ and is free of the notional amounts, it represents the portion of the population basis risk that exists no matter what the notional amounts of the

Table 4.1: A summary of the information about the variance components $V_1(t), \ldots, V_5(t)$. The definitions of $\Psi_t^{(P_H)}$, $\vec{G}_t^{(P_H)}$ and $\Gamma_t^{(P_H)}$ are provided in equations (4.8), (4.10) and (4.9), respectively.

Component	Relation with $\vec{N}_t^{(P_H)}$	Associated hidden states	Matrix notation
$V_1(t)$	Constant	$\vec{\alpha}_t^c$	Not applicable
$V_2(t)$	Positive and quadratic	$\vec{\alpha}_t^c$	$(\vec{N}_t^{(P_H)})'\Psi_t^{(P_H)}\vec{N}_t^{(P_H)}$
$V_3(t)$	Negative and linear	$\vec{\alpha}_t^c$	$-2(\vec{G}_t^{(P_H)})'\vec{N}_t^{(P_H)}$
$V_4(t)$	Constant	$\vec{\alpha}_t^{(P_L)}$	Not applicable
$V_5(t)$	Positive and quadratic	$\vec{\alpha}_t^{(P_H)}$	$(\vec{N}_t^{(P_H)})'\Gamma_t^{(P_H)}\vec{N}_t^{(P_H)}$

q-forwards are. On the other hand, $V_5(t)$ increases with the q-forwards' notional amounts in a quadratic manner. This relationship is a result of the fact that the hedger is subject to more population basis risk as it acquires larger notional amounts of q-forwards that are linked to a population which is different from its own population of individuals.

Overall, as the hedger acquires q-forwards, $V_3(t)$ reduces but $V_2(t)$ and $V_5(t)$ increase. If the reduction in $V_3(t)$ outweighs the increase in $V_2(t)$ and $V_5(t)$, then the total variance of the hedged position reduces. Furthermore, because the effects of $V_3(t)$ and the sum of $V_4(t)$ and $V_5(t)$ are offsetting, there exists an optimal vector of notional amounts which would minimize the variance of the hedged position. In Table 4.1 we summarize how each variance component is related to the q-forwards' notional amounts and the hidden states. These relationships form the basis for the derivation of the optimal notional amounts in Section 4.3.3 and the further analysis of population basis risk in Section 4.4.

4.3.3 Deriving the Hedging Strategies

We now derive the optimal notional amounts that minimize $\text{Var}(l(t)-\sum_{j=1}^{m} N_j^{(P_H)}(t)h_j^{(P_H)}(t))$, the approximated variance of the values of the hedged position. To facilitate the derivation, we first rewrite $V_2(t)$, $V_3(t)$ and $V_5(t)$ in matrix forms. For $V_2(t)$ and $V_5(t)$, which are quadratically related to the notional amounts, we have

$$V_2(t) = (\vec{N}_t^{(P_H)})'\Psi_t^{(P_H)}\vec{N}_t^{(P_H)},$$

$$V_5(t) = (\vec{N}_t^{(P_H)})'\Gamma_t^{(P_H)}\vec{N}_t^{(P_H)},$$

where $\Psi_t^{(P_H)}$ and $\Gamma_t^{(P_H)}$ are m-by-m square matrices, with the (i,j)th element being

$$\Psi_{i,j|t}^{(P_H)} = \left(\frac{\partial H_i^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_i|t}^c}\right)' \mathrm{Cov}(\vec{\alpha}_{t+T_i|t}^c, \vec{\alpha}_{t+T_j|t}^c) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^c} \tag{4.8}$$

and

$$\Gamma_{i,j|t}^{(P_H)} = \left(\frac{\partial H_i^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_i|t}^{(P_H)}}\right)' \mathrm{Cov}(\vec{\alpha}_{t+T_i|t}^{(P_H)}, \vec{\alpha}_{t+T_j|t}^{(P_H)}) \frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^{(P_H)}}, \tag{4.9}$$

respectively. For $V_3(t)$, which is linearly related to the notional amounts, we have

$$V_3(t) = -2(\vec{G}_t^{(P_H)})' \vec{N}_t^{(P_H)},$$

where $\vec{G}_t^{(P_H)}$ is an m-by-1 vector with the jth element $G_j^{(P_H)}(t)$ being

$$G_j^{(P_H)}(t) = \sum_{s=t+1}^{t_0+T} \left(\frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^c}\right)' \mathrm{Cov}(\vec{\alpha}_{t+T_j|t}^c, \vec{\alpha}_{s|t}^c) \frac{\partial L(t)}{\partial \vec{\alpha}_{s|t}^c}. \tag{4.10}$$

We can then express the variance to be minimized as

$$\mathrm{Var}\left(l(t) - \sum_{j=1}^m N_j^{(P_H)}(t) h_j^{(P_H)}(t)\right) = (\vec{N}_t^{(P_H)})'(\Psi_t^{(P_H)} + \Gamma_t^{(P_H)})\vec{N}_t^{(P_H)} - 2(\vec{G}_t^{(P_H)})'\vec{N}_t^{(P_H)} + C(t), \tag{4.11}$$

where $C(t) = V_1(t) + V_4(t)$ is a constant that is free of the notional amounts.

To derive the optimal notional amounts, we take partial derivative of $\mathrm{Var}(l(t) - \sum_{j=1}^m N_j^{(P_H)}(t) h_j^{(P_H)}(t))$ with respect to $\vec{N}_t^{(P_H)}$, which gives

$$\frac{\partial \mathrm{Var}(l(t) - \sum_{j=1}^m N_j^{(P_H)}(t) h_j^{(P_H)}(t))}{\partial \vec{N}_t^{(P_H)}} = 2(\Psi_t^{(P_H)} + \Gamma_t^{(P_H)})\vec{N}_t^{(P_H)} - 2\vec{G}_t^{(P_H)}.$$

The optimal vector of notional amounts $\vec{\hat{N}}_t^{(P_H)}$ that minimizes $\mathrm{Var}(l(t) - \sum_{j=1}^m N_j^{(P_H)}(t) h_j^{(P_H)}(t))$

can be obtained readily by setting the partial derivative to zero:

$$\begin{aligned}\left(\Psi_t^{(P_H)} + \Gamma_t^{(P_H)}\right)\vec{\hat{N}}_t^{(P_H)} &= \vec{G}_t^{(P_H)} \\ \vec{\hat{N}}_t^{(P_H)} &= \left(\Psi_t^{(P_H)} + \Gamma_t^{(P_H)}\right)^{-1}\vec{G}_t^{(P_H)}.\end{aligned} \tag{4.12}$$

The second step holds because, as discussed in Chapter 3, $\Psi_t^{(P_H)}$ and $\Gamma_t^{(P_H)}$ are positive-definite matrices. Finally, the minimized value of $\mathrm{Var}(l(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t)h_j^{(P_H)}(t))$ can be obtained by plugging equation (4.12) into equation (4.11), which gives

$$\min_{N_1^{(P_H)}(t),\ldots,N_m^{(P_H)}(t)} \left(\mathrm{Var}(l(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t)h_j^{(P_H)}(t))\right) = C(t) - (\vec{G}_t^{(P_H)})'\left(\Psi_t^{(P_H)} + \Gamma_t^{(P_H)}\right)^{-1}\vec{G}_t^{(P_H)}. \tag{4.13}$$

Equations (4.12) and (4.13) involve $\mathrm{Cov}(\vec{\alpha}_{s|t}^{c}, \vec{\alpha}_{u|t}^{c})$, $\mathrm{Cov}(\vec{\alpha}_{s|t}^{(P_L)}, \vec{\alpha}_{u|t}^{(P_L)})$ and $\mathrm{Cov}(\vec{\alpha}_{s|t}^{(P_H)}, \vec{\alpha}_{u|t}^{(P_H)})$ for some $s, u > t$. To compute these quantities, we first calculate the covariance matrix of $\vec{\alpha}_{s|t}$ and $\vec{\alpha}_{u|t}$ for $s, u > t$, given the information up to and including time t, as

$$\Xi_{s,u|t} := \mathrm{Cov}(\vec{\alpha}_{s|t}, \vec{\alpha}_{u|t}) = \begin{cases} D^{|s-u|}(Q + DQD' + \cdots D^{u-(t+1)}Q(D^{u-(t+1)})') & , s > u \\ (Q + DQD' + \cdots D^{s-(t+1)}Q(D^{s-(t+1)})')(D^{|s-u|})' & , s < u \\ Q + DQD' + \cdots D^{u-(t+1)}Q(D^{u-(t+1)})' & , s = u \end{cases}.$$

Then $\Xi_{s,u|t}$ is decomposed into $(n_p + 1)$ block matrices as follows:

$$\Xi_{s,u|t} = \begin{pmatrix} \mathrm{Cov}(\vec{\alpha}_{s|t}^{c}, \vec{\alpha}_{u|t}^{c}) & 0 & 0 & \cdots & 0 \\ 0 & \mathrm{Cov}(\vec{\alpha}_{s|t}^{(1)}, \vec{\alpha}_{u|t}^{(1)}) & 0 & \cdots & 0 \\ 0 & 0 & \mathrm{Cov}(\vec{\alpha}_{s|t}^{(2)}, \vec{\alpha}_{u|t}^{(2)}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \mathrm{Cov}(\vec{\alpha}_{s|t}^{(n_p)}, \vec{\alpha}_{u|t}^{(n_p)}) \end{pmatrix}.$$

Finally, $\Xi_{s,u|t}^{c} = \mathrm{Cov}(\vec{\alpha}_{s|t}^{c}, \vec{\alpha}_{u|t}^{c})$, $\Xi_{s,u|t}^{(P_L)} = \mathrm{Cov}(\vec{\alpha}_{s|t}^{(P_L)}, \vec{\alpha}_{u|t}^{(P_L)})$ and $\Xi_{s,u|t}^{(P_H)} = \mathrm{Cov}(\vec{\alpha}_{s|t}^{(P_H)}, \vec{\alpha}_{u|t}^{(P_H)})$

can be obtained respectively from the corresponding block matrices in $\Xi_{s,u|t}$.

Equations (4.12) and (4.13) also encompass the partial derivatives of $L(t)$ and $H_j^{(P_H)}(t)$ with respect to the common states and the population-specific states. The partial derivatives of $L(t)$ with respect to $\vec{\alpha}^c_{s|t}$ and $\vec{\alpha}^{(P_L)}_{s|t}$, evaluated at $\vec{\alpha}^c_{s|t} = \vec{\hat{\alpha}}^c_{s|t}$ and $\vec{\alpha}^{(P_L)}_{s|t} = \vec{\hat{\alpha}}^{(P_L)}_{s|t}$, can be computed as

$$\begin{aligned}
\frac{\partial L(t)}{\partial \vec{\alpha}^c_{s|t}} &= \sum_{u=s-t}^{T-(t-t_0)} e^{-ru} \left(\frac{\partial g^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\frac{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial \vec{\alpha}^c_{s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\prod_{\substack{v=t+1 \\ v\neq s}}^{t+u} \hat{g}^{(P_L)}_{x_0+v-t_0-1,v|t} \right) \\
&= \sum_{u=s-t}^{T-(t-t_0)} e^{-ru} \vec{b}^c_{x_0+s-t_0-1} \left(\frac{\partial g^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\prod_{\substack{v=t+1 \\ v\neq s}}^{t+u} \hat{g}^{(P_L)}_{x_0+v-t_0-1,v|t} \right)
\end{aligned} \tag{4.14}$$

and

$$\begin{aligned}
\frac{\partial L(t)}{\partial \vec{\alpha}^{(P_L)}_{s|t}} &= \sum_{u=s-t}^{T-(t-t_0)} e^{-ru} \left(\frac{\partial g^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\frac{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial \vec{\alpha}^{(P_L)}_{s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\prod_{\substack{v=t+1 \\ v\neq s}}^{t+u} \hat{g}^{(P_L)}_{x_0+v-t_0-1,v|t} \right) \\
&= \sum_{u=s-t}^{T-(t-t_0)} e^{-ru} \vec{b}^{(P_L)}_{x_0+s-t_0-1} \left(\frac{\partial g^{(P_L)}_{x_0+s-t_0-1,s|t}}{\partial y^{(P_L)}_{x_0+s-t_0-1,s|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\prod_{\substack{v=t+1 \\ v\neq s}}^{t+u} \hat{g}^{(P_L)}_{x_0+v-t_0-1,v|t} \right),
\end{aligned} \tag{4.15}$$

respectively, where $\vec{b}^c_x$ and $\vec{b}^{(P_L)}_x$ are defined in equation (4.3), and $\hat{g}^{(P_L)}_{x,s|t}$ is the value of $g^{(P_L)}_{x,s|t}$ evaluated at $\vec{\alpha}^c_{s|t} = \vec{\hat{\alpha}}^c_{s|t}$ and $\vec{\alpha}^{(P_L)}_{s|t} = \vec{\hat{\alpha}}^{(P_L)}_{s|t}$. Similarly, the partial derivatives of $H_j^{(P_H)}(t)$ with respect to $\vec{\alpha}^c_{t+T_j|t}$ and $\vec{\alpha}^{(P_H)}_{t+T_j|t}$, evaluated at $\vec{\alpha}^c_{t+T_j|t} = \vec{\hat{\alpha}}^c_{t+T_j|t}$ and $\vec{\alpha}^{(P_H)}_{t+T_j|t} = \vec{\hat{\alpha}}^{(P_H)}_{t+T_j|t}$, can be calculated as

$$\begin{aligned}
\frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}^c_{t+T_j|t}} &= e^{-rT_j} \left(\frac{\partial g^{(P_H)}_{x_j,t+T_j|t}}{\partial y^{(P_H)}_{x_j,t+T_j|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \left(\frac{\partial y^{(P_H)}_{x_j,t+T_j|t}}{\partial \vec{\alpha}^c_{t+T_j|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right) \\
&= e^{-rT_j} \vec{b}^c_{x_j} \left(\frac{\partial g^{(P_H)}_{x_j,t+T_j|t}}{\partial y^{(P_H)}_{x_j,t+T_j|t}} \bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} \right)
\end{aligned} \tag{4.16}$$

and

$$\begin{aligned}\frac{\partial H_j^{(P_H)}(t)}{\partial \vec{\alpha}_{t+T_j|t}^{(P_H)}} &= e^{-rT_j}\left(\frac{\partial g_{x_j,t+T_j|t}^{(P_H)}}{\partial y_{x_j,t+T_j|t}^{(P_H)}}\bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}}\right)\left(\frac{\partial y_{x_j,t+T_j|t}^{(P_H)}}{\partial \vec{\alpha}_{t+T_j|t}^{(P_H)}}\bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}}\right) \\ &= e^{-rT_j}\vec{b}_{x_j}^{(P_H)}\left(\frac{\partial g_{x_j,t+T_j|t}^{(P_H)}}{\partial y_{x_j,t+T_j|t}^{(P_H)}}\bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}}\right),\end{aligned} \tag{4.17}$$

respectively, where $\vec{b}_x^c$ and $\vec{b}_x^{(P_H)}$ are defined in equation (4.3).

The calculation of $\partial g_{x,s|t}^{(p)}/\partial y_{x,s|t}^{(p)}$ in equations (4.14), (4.15), (4.16) and (4.17) depends on the specification of the assumed model. For the ACF model (and other models that are built for central death rates), we have

$$\frac{\partial g_{x,s|t}^{(p)}}{\partial y_{x,s|t}^{(p)}}\bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} = -\exp(\hat{y}_{x,s|t}^{(p)})\exp(-\exp(\hat{y}_{x,s|t}^{(p)})) = -\hat{m}_{x,s|t}^{(p)}\exp(-\hat{m}_{x,s|t}^{(p)}), \tag{4.18}$$

where $\hat{m}_{x,s|t}^{(p)}$ represents the value of $m_{x,s|t}^{(p)}$ evaluated at $\vec{\alpha}_{s|t}^c = \vec{\hat{\alpha}}_{s|t}^c$ and $\vec{\alpha}_{s|t}^{(p)} = \vec{\hat{\alpha}}_{s|t}^{(p)}$. For the M-CBD model (and other models that are built for single-year conditional death probabilities), we have

$$\frac{\partial g_{x,s|t}^{(p)}}{\partial y_{x,s|t}^{(p)}}\bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}} = -\frac{\exp(\hat{y}_{x,s|t}^{(p)})}{(1+\exp(\hat{y}_{x,s|t}^{(p)}))^2} = -\hat{q}_{x,s|t}^{(p)}\left(1-\hat{q}_{x,s|t}^{(p)}\right), \tag{4.19}$$

where $\hat{q}_{x,s|t}^{(p)}$ represents the value of $q_{x,s|t}^{(p)}$ evaluated at $\vec{\alpha}_{s|t}^c = \vec{\hat{\alpha}}_{s|t}^c$ and $\vec{\alpha}_{s|t}^{(p)} = \vec{\hat{\alpha}}_{s|t}^{(p)}$.

4.3.4 Evaluation of Hedge Effectiveness

We evaluate hedge effectiveness by measuring the proportion of variance reduced. In the absence of any longevity hedge, the time-t_0 (random) value of the annuity liabilities is $L(t_0)$. If a static hedge with m q-forwards is established at $t = t_0$, then the hedger would receive offsetting cash flows from the q-forwards at $t = t_0 + T_1, \ldots, t_0 + T_m$, which collectively have a time-t_0 (random) value of $\sum_{j=1}^m N_j^{(P_H)}(t_0)H_j^{(P_H)}(t_0)$. If the hedge is effective, then the offsetting cash flows would result in a hedged position that has a small variance. Using this

reasoning, we assess the hedge effectiveness of a static hedge with the following metric:

$$HE = 1 - \frac{\text{Var}\left(L(t_0) - \sum_{j=1}^{m} N_j^{(P_H)}(t_0) H_j^{(P_H)}(t_0)\right)}{\text{Var}(L(t_0))}.$$

The value of HE is close to 1 if the longevity hedge is effective, and close to 0 if otherwise.

Using the approximation technique and the variance decomposition discussed in Section 4.3.2, we can approximate HE as

$$\widehat{HE} = 1 - \frac{\text{Var}\left(l(t_0) - \sum_{j=1}^{m} N_j^{(P_H)}(t_0) h_j^{(P_H)}(t_0)\right)}{\text{Var}(l(t_0))} = 1 - \frac{V_1(t_0) + V_2(t_0) + V_3(t_0) + V_4(t_0) + V_5(t_0)}{V_1(t_0) + V_4(t_0)}.$$

As shown in Section 4.3.2, all components (i.e., the covariance terms and the partial derivatives) in $V_1(t_0)$, $V_2(t_0)$, $V_3(t_0)$, $V_4(t_0)$ and $V_5(t_0)$ can be analytically computed. Hence, with minimal computational effort, we can calculate $\widehat{HE}$ for different combinations of notional amounts and derive an empirical relationship between $\widehat{HE}$ and $N_1^{(P_H)}, \ldots, N_m^{(P_H)}$. This feature may be considered as an advantage over some of the existing longevity hedging methods (Cairns, 2011, 2013; Cairns et al., 2014; Coughlan et al., 2011; Li and Hardy, 2011), in which a simulation is required to calculate the hedge effectiveness for each combination of notional amounts.

For a dynamic hedge, the cash flows from the hedge portfolio are more complicated because the hedger does not hold the hedging instruments to maturity. In constructing a dynamic hedge, we assume that at time t_0 the hedger acquires m freshly launched q-forwards which have reference ages $x_1, \ldots, x_m$ and times-to-maturity $T_1, \ldots, T_m$. Then, at each time point $t = t_0 + 1, \ldots, t_0 + T - 1$, the hedger closes out all of the q-forwards in the existing hedge portfolio, and acquires m freshly launched q-forwards which also have reference ages $x_1, \ldots, x_m$ and times-to-maturity $T_1, \ldots, T_m$. Finally, at $t = T$, all of the q-forwards in the hedge portfolio are closed out. We let $PCF(t)$ be time-t_0 value of the net cash flow that is incurred when the hedger adjusts the hedge portfolio at time t. Given the assumptions we made, we can express $PCF(t)$ as follows:

$$PCF(t) = S_{x_0,t_0,t-1}^{(P_H)} \sum_{j=1}^{m} N_j^{(P_H)}(t-1) e^{-r(t-t_0-1+T_j)} (\text{E}(q_{x_j,t-1+T_j|t-1}^{(P_H)}) - \text{E}(q_{x_j,t-1+T_j|t}^{(P_H)})) \quad (4.20)$$

for $t = t_0 + 1, \ldots, t_0 + T$, where

$$S^{(P_H)}_{x_0,t_0,t-1} = \begin{cases} 1, & t = t_0 + 1 \\ \prod_{s=t_0+1}^{t-1}(1 - q^{(P_H)}_{x_0+s-t_0-1,s|t-1}), & t = t_0 + 2, \ldots, t_0 + T \end{cases}$$

represents the probability that an individual from the hedger's population, who is aged x_0 at time t_0, survives to time $t - 1$. The effectiveness for the dynamic hedge can then be assessed by the following formula:

$$HE = 1 - \frac{\text{Var}\left(L(t_0) - \sum_{t=t_0+1}^{t_0+T} PCF(t)|\mathcal{F}_{t_0}\right)}{\text{Var}(L(t_0)|\mathcal{F}_{t_0})}.$$

4.4 Analyzing Population Basis Risk

The goal of this section is to investigate how the variance decomposition presented in Section 4.3.2 can help us quantify population basis risk. We begin this section by studying the hedger's risk exposure in a situation when population basis risk is assumed to be absent. Then, we examine the effect of population basis risk by observing how the hedger's risk exposure would change when the assumption of no population basis risk is relaxed. Finally, we develop a quantity called 'standardized basis risk profile,' which allows hedgers to compare the levels of population basis risk arising from different reference populations. Throughout this section, diagrams that aid us to understand population basis risk are presented.

4.4.1 The Hedger's Risk Exposure when Population Basis Risk is Absent

Let us first consider a simplified scenario in which population basis risk is assumed to be absent. Under our set-up, population basis risk arises solely from the population-specific states, so the simplified scenario can be interpreted to mean the situation when the population-specific states are not stochastic. To construct the simplified scenario, we set $Q^{(p)} = 0$ for $p = 1, 2, \ldots, n_p$, which implies the population-specific states would remain constant; that is, $\vec{\alpha}^{(p)}_{s|t_0} \equiv \vec{\alpha}^{(p)}_{t_0|t_0}$, for $p = 1, \ldots, n_p$ and $s \geq t_0$. Consequently, $\text{Cov}(\vec{\alpha}^{(P_H)}_{s|t}, \vec{\alpha}^{(P_H)}_{u|t}) = 0$ for all $s, u > t$, and hence $\Gamma^{(P_H)}_t$, $V_4(t)$ and $V_5(t)$ would become zero. In the simplified scenario, the total longevity risk that the hedger is exposed to is $V_1(t) + V_2(t) + V_3(t)$.

When population basis risk is hypothetically absent, the choice of the q-forwards' reference population P_H should not affect the optimal hedge effectiveness. In other words, the minimized variance specified in equation (4.13) should not depend on P_H. This fact can be verified mathematically as follows. Let us rewrite the square matrix $\Psi_t^{(P_H)}$ in equation (4.11) as a product of three square matrices,

$$\Psi_t^{(P_H)} = \Lambda_t^{(P_H)} Z_t^c \Lambda_t^{(P_H)}, \tag{4.21}$$

where $\Lambda_t^{(P_H)}$ is an m-by-m diagonal matrix with the jth diagonal element $\lambda_j^{(P_H)}(t)$ being

$$\lambda_j^{(P_H)}(t) = e^{-rT_j} \frac{\partial g_{x_j,t+T_j|t}^{(P_H)}}{\partial y_{x_j,t+T_j|t}^{(P_H)}} \Bigg|_{\vec{\alpha}_{s|t}=\vec{\hat{\alpha}}_{s|t}}, \tag{4.22}$$

and Z_t^c is another m-by-m symmetric matrix with the (i,j)th element being

$$Z_{i,j}^c(t) = (\vec{b}_{x_i}^c)' \Xi_{t+T_i,t+T_j|t}^c \vec{b}_{x_j}^c. \tag{4.23}$$

Similarly, we rewrite $\vec{G}_t^{(P_H)}$ in equation (4.11) as a product of a square matrix and a vector,

$$\vec{G}_t^{(P_H)} = \Lambda_t^{(P_H)} \vec{G}_t^c, \tag{4.24}$$

where $\vec{G}_t^c = (G_1^c(t), \ldots, G_m^c(t))'$ is an m-by-1 vector with the jth element $G_j^c(t)$ being

$$G_j^c(t) = \sum_{s=t+1}^{t_0+T} (\vec{b}_{x_j}^c)' \Xi_{t+T_j,s|t}^c \frac{\partial L(t)}{\partial \vec{\alpha}_{s|t}}.$$

The definition of $\vec{b}_x^c$ is provided in equation (4.3). Note that neither Z_t^c nor $\vec{G}_t^c$ depends on P_H. Substituting $\Gamma_t^{(P_H)} = 0$ and equations (4.21) and (4.24) into equation (4.13), we obtain

$$\min_{N_1^{(P_H)}(t),\ldots,N_m^{(P_H)}(t)} \left(\mathrm{Var}(l(t) - \sum_{j=1}^m N_j^{(P_H)}(t) h_j^{(P_H)}(t)) \right) = C(t) - (\vec{G}_t^c)'(Z_t^c)^{-1} \vec{G}_t^c,$$

which is free of P_H.

Nevertheless, even if population basis risk is absent, the optimal notional amounts do

depend on the q-forwards' reference population. The dependence of $\vec{\hat{N}}_t^{(P_H)}$ on P_H can be seen by substituting $\Gamma_t^{(P_H)} = 0$ and equations (4.21) and (4.24) into equation (4.12), which gives

$$\vec{\hat{N}}_t^{(P_H)} = (\Lambda_t^{(P_H)})^{-1}(Z_t^c)^{-1}\vec{G}_t^c. \tag{4.25}$$

The dependence of $\vec{\hat{N}}_t^{(P_H)}$ on P_H lies in the diagonal matrix $\Lambda_t^{(P_H)}$, which, according to equations (4.18) and (4.19), contains either $\hat{m}_{x,s|t}^{(P_H)}$ or $\hat{q}_{x,s|t}^{(P_H)}$. This diagonal matrix is related to P_H, because different reference populations may have different expected levels of mortality even after controlling for age and time.

To facilitate comparison, we define a quantity called 'standardized notional amount,' which takes the expected level of mortality into account. Let

$$\vec{\mathcal{N}}_t^{(P_H)} = (\mathcal{N}_1^{(P_H)}(t), \ldots, \mathcal{N}_m^{(P_H)}(t))' = \Lambda_t^{(P_H)}\vec{N}_t^{(P_H)}$$

be the vector of standardized notional amounts. It follows from equation (4.25) that the optimal standardized notional amounts $\vec{\hat{\mathcal{N}}}_t^{(P_H)} = \Lambda_t^{(P_H)}\vec{\hat{N}}_t^{(P_H)} = (Z_t^c)^{-1}\vec{G}_t^c$ are free of P_H. Furthermore, when written in terms of $\vec{\mathcal{N}}_t^{(P_H)}$, both $V_2(t) = (\vec{\mathcal{N}}_t^{(P_H)})'Z_t^c\vec{\mathcal{N}}_t^{(P_H)}$ and $V_3(t) = -2(\vec{G}_t^c)'\vec{\mathcal{N}}_t^{(P_H)}$ have coefficients that do not depend on P_H.

The arguments above can be demonstrated graphically. Let us consider the case when $m = 1$ q-forward is used and three hypothetical reference populations, p_1, p_2 and p_3, are available. The left panel of Figure 4.1 shows, for each reference population, the theoretical pattern of $V_1(t)+V_2(t)+V_3(t)$ (i.e., the variance of the hedged position in the absence of any population basis risk) as a function of the (non-standardized) notional amount $N_1^{(P_H)}(t)$. The following four properties are noted: (i) the curves have a parabolic shape, because $V_1(t) + V_2(t) + V_3(t)$ is a quadratic function of $N_1^{(P_H)}(t)$; (ii) the curves have a common vertical intercept of $V_1(t)$, because the variance must be $V_1(t)$ if the annuity liabilities are left unhedged; (iii) the curves have the same minimum value, because the choice of P_H should have no effect on the optimal hedge effectiveness when population basis risk is absent; (iv) the curves have different axes of symmetry, owing to the effect of $\Lambda_t^{(P_H)}$. The right panel of Figure 4.1 considers the standardized notional amount $\mathcal{N}_1^{(P_H)}(t)$ instead. As expected, properties (i) to (iii) still hold, but the three curves are overlapping. The overlap is because when $V_1(t) + V_2(t) + V_3(t)$ is expressed as a function of $\mathcal{N}_1^{(P_H)}(t)$, its coefficients are no longer dependent on P_H.

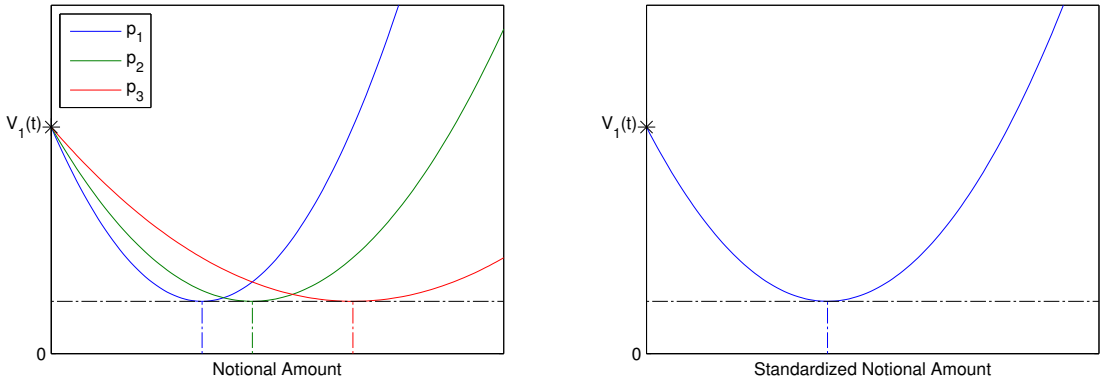


Figure 4.1: The theoretical patterns of $V_1(t)+V_2(t)+V_3(t)$ as functions of the (non-standardized) notional amount (the left panel) and the standardized notional amount (the right panel). It is assumed that one q-forward is used and that the available reference populations are p_1, p_2 and p_3.

4.4.2 The Hedger's Risk Exposure when Population Basis Risk is Present

When population basis risk is present, $\Gamma_t^{(P_H)}$ becomes non-zero and the hedger's total exposure to longevity risk is $V_1(t)+V_2(t)+V_3(t)+V_4(t)+V_5(t)$. To have a deeper understanding of the hedger's risk exposure, we rewrite $\Gamma_t^{(P_H)}$ as a product of three square matrices,

$$\Gamma_t^{(P_H)} = \Lambda_t^{(P_H)} Z_t^{(P_H)} \Lambda_t^{(P_H)},$$

where $\Lambda_t^{(P_H)}$ is specified in equation (4.22) and $Z_t^{(P_H)}$ is an m-by-m symmetric matrix with the (i,j)th element being

$$Z_{i,j}^{(P_H)}(t) = (\vec{b}_{x_i}^{(P_H)})' \Xi_{t+T_i,t+T_j|t}^{(P_H)} \vec{b}_{x_j}^{(P_H)}. \tag{4.26}$$

The definition of $\vec{b}_x^{(P_H)}$ is provided in equation (4.3). It immediately follows from equation (4.12) that the optimal standardized notional amounts are given by

$$\vec{\mathcal{N}}_t^{(P_H)} = \Lambda_t^{(P_H)} \vec{\mathcal{N}}_t^{(P_H)} = (Z_t^c + Z_t^{(P_H)})^{-1} \vec{G}_t^c. \tag{4.27}$$

Also, it follows from equation (4.13) that the minimized variance can be expressed as

$$\min_{N_1^{(P_H)}(t),\ldots,N_m^{(P_H)}(t)} \left(\text{Var}(l(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t) h_j^{(P_H)}(t)) \right) = C(t) - (\vec{G}_t^c)'(Z_t^c + Z_t^{(P_H)})^{-1}\vec{G}_t^c. \tag{4.28}$$

When population basis risk is present, $Z_t^{(P_H)}$, $\vec{\mathcal{N}}_t^{(P_H)}$ and the minimized variance are all related to the q-forwards' reference population P_H. Moreover, because $C(t)$, $\vec{G}_t^c$ and Z_t^c do not depend on P_H, the dependence of $\vec{\mathcal{N}}_t^{(P_H)}$ and the minimized variance on P_H lies exclusively in $Z_t^{(P_H)}$. Therefore, $Z_t^{(P_H)}$ should contain all information concerning the population basis risk that arises from the difference in the mortality improvements between the reference population P_H and the hedger's own population of individuals.

Let us consider the case when $m = 1$ q-forward is used. In this case, $Z_t^{(P_H)}$ reduces to a scalar and equals

$$Z_{1,1}^{(P_H)}(t) = (\vec{b}_{x_1}^{(P_H)})' \Xi_{t+T_1,t+T_1|t}^{(P_H)} \vec{b}_{x_1}^{(P_H)} := BRP(x_1, T_1, P_H).$$

We call $BRP(x_1, T_1, P_H)$ the 'standardized basis risk profile' for a q-forward with reference age x_1, time-to-maturity T_1 and reference population P_H. This quantity is non-negative, because it can be regarded as the variance of $(\vec{b}_{x_1}^{(P_H)})' \vec{\alpha}_{t+T_1|t}^{(P_H)}$. It can be observed from equation (4.28) that when $m = 1$, the minimized variance is an increasing function of $BRP(x_1, T_1, P_H)$. A smaller value of $BRP(x_1, T_1, P_H)$ thus represents a more favourable optimal hedging performance. Moreover, according to equation (4.27), when $m = 1$ the optimal standardized notional amount is a decreasing function in $BRP(x_1, T_1, P_H)$. What this means is that in the extreme case when $BRP(x_1, T_1, P_H)$ is very high, it is optimal not to hedge with the q-forward with parameters x_1, T_1 and P_H.

The connection between $BRP(x_1, T_1, P_H)$ and population basis risk can be understood from another angle. According to the decomposition of variance presented in Section 4.3.2, population basis risk is represented by two components, $V_4(t)$ and $V_5(t)$, which are respectively related to the idiosyncratic features of the hedger's own population and the q-forwards' reference population. Also, while $V_4(t)$ is a constant, $V_5(t)$ is a quadratic function of the q-forwards' notional amounts, reflecting the fact that the hedger is exposed to more risk associated with the idiosyncratic features of the q-forwards' reference population as the notional amounts become larger. As a fact, when $m = 1$,

$$V_5(t) = BRP(x_1, T_1, P_H) \times (\mathcal{N}_1^{(P_H)}(t))^2,$$

which means that $BRP(x_1, T_1, P_H)$ can be understood as the speed at which population basis risk grows when the (standardized) notional amount of the q-forward with parameters x_1, T_1 and P_H increases. It is clear that other things equal, a hedger should choose a q-forward with the lowest value of $BRP(x_1, T_1, P_H)$.

In practice, the standardized basis risk profile can aid hedgers to select the most appropriate reference population out of the reference populations available in the market. While the methods proposed previously by Cairns (2011, 2013), Cairns et al. (2014), Coughlan et al. (2011) and Li and Hardy (2011) may also be used for the purpose of reference population selection, they generally require a lot more computational effort because they involve multiple steps: (i) for each candidate reference population, calibrate the hedge (i.e., select the optimal notional amounts); (ii) simulate realizations of future mortality and calculate the hedge effectiveness resulting from each calibrated hedge; (iii) choose the reference population that leads to the best hedge effectiveness. In contrast, the standardized basis risk profile can be computed analytically, and the comparison between different standardized basis risk profiles can be made without calibrating the associated hedges in advance.

We now revisit the previously made arguments from a geometrical viewpoint. Let us continue to assume that the hedge portfolio contains $m = 1$ q-forward. The upper-left panel of Figure 4.2 shows the theoretical pattern of $V_4(t) + V_5(t)$ (the portion of variance arising from population basis risk) as a function of the standardized notional amount $\mathcal{N}_1^{(P_H)}(t)$. Since $V_4(t)$ is a constant and $V_5(t)$ increases quadratically with $\mathcal{N}_1^{(P_H)}(t)$, the pattern is a parabola with $V_4(t)$ as the vertical intercept and $\mathcal{N}_1^{(P_H)}(t) = 0$ as the axis of symmetry.

The upper-right panel of Figure 4.2 compares the theoretical patterns of $V_1(t) + V_2(t) + V_3(t)$ (the total variance in the absence of population basis risk) and $V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t)$ (the total variance including population basis risk). As population basis risk is being incorporated, $V_4(t) + V_5(t)$ is superimposed onto $V_1(t) + V_2(t) + V_3(t)$, leading the pattern of the total variance to shift upwards and leftwards. The upward shift reflects, for every $\mathcal{N}_1^{(P_H)}(t) > 0$, the loss in hedge effectiveness that is owing to population basis risk. The leftward shift highlights the fact that the optimal (standardized) notional amount would be smaller when population basis risk is taken into consideration.

The lower-left panel of Figure 4.2 plots $V_4(t)+V_5(t)$ against $\mathcal{N}_1^{(P_H)}(t)$ for three q-forwards with the same reference age x_1 and time-to-maturity T_1 but different reference populations ($P_H = p_1, p_2, p_3$). We further assume that $BRP(x_1, T_1, p_3) > BRP(x_1, T_1, p_1) > BRP(x_1, T_1, p_2)$. All three curves intersect the vertical axis at $V_4(t)$, because $V_5(t) = 0$ when $\mathcal{N}_1^{(P_H)}(t) = 0$. However, they have different curvatures, which are determined by their corresponding standardized population basis risk profiles. Finally, in lower-right panel of Figure 4.2 we show the theoretical patterns of $V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t)$ for the

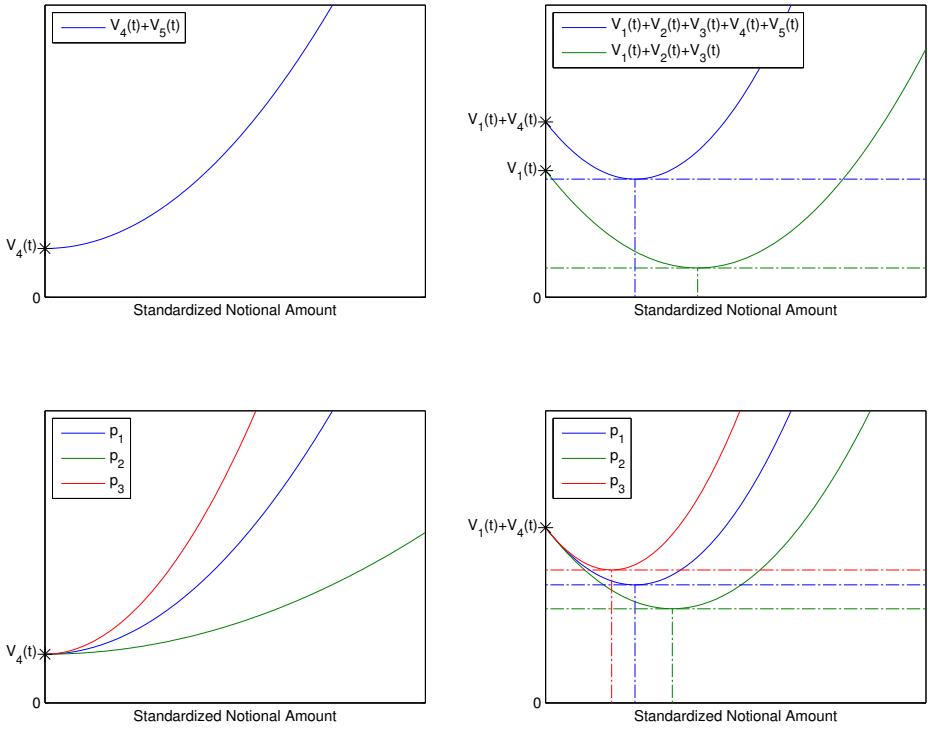


Figure 4.2: The theoretical relationships between different combinations of variance components and the standardized notional amount of the q-forward in a single-instrument hedge portfolio;
upper-left: $V_4(t) + V_5(t)$;
upper-right: $V_1(t) + V_2(t) + V_3(t)$ and $V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t)$;
lower-left: $V_4(t) + V_5(t)$ for hypothetical reference populations p_1, p_2, p_3;
lower-right: $V_1(t) + V_2(t) + V_3(t) + V_4(t) + V_5(t)$ for hypothetical reference populations p_1, p_2, p_3.

three q-forwards that are linked to reference populations p_1, p_2 and p_3. Among the five variance components, $V_5(t)$ is the only one that is related to P_H and hence the differences between the shapes of the curves are due entirely to the differences in $V_5(t)$ or equivalently $BRP(x_1, T_1, P_H)$. The curve with the lowest vertex (which means the highest hedge effectiveness) is the one that corresponds to the q-forward with the smallest standardized population basis risk profile.

4.5 A Numerical Illustration

In this section, we illustrate the proposed methods with some real mortality data. We begin this section by stating the general assumptions made. We then present the estimated multi-population mortality model on which the illustrative longevity hedges are based. Finally, the results of the illustrative static and dynamic longevity hedges are discussed in turn.

4.5.1 Assumptions

The following assumptions are made throughout the rest of this section.

- The liability being hedged is a portfolio of 30-year temporary life annuity immediate contracts that are sold to persons aged 60 at time t_0; that is $x_0 = 60$ and $T = 30$.
- The annuitants' mortality is exactly the same as that of Canadian males.
- There is no small sample risk.
- The hedge portfolio is composed of only $m = 1$ q-forward, which has a reference age of $x_1 = 60$ and a time-to-maturity (from the launch date) of $T_1 = 10$.
- In the market, q-forwards that are linked to the male populations of the following four countries are available and liquidly traded: the US, England and Wales, the Netherlands and West Germany. We make this assumption because the tradable LifeMetrics Indexes offered by the LLMA are linked to each of these four national populations. There is no q-forward linked to the population of Canadian males.
- The continuously compounded interest rate (r) is assumed to be 0.01 per annum.
- There is no transaction cost.

4.5.2 The Multi-Population Mortality Model Used

Given the assumptions made, we require a mortality model for $n_p = 5$ populations. The model we use is the ACF model discussed in Section 4.2.2. We estimate the model to the historical mortality data of Canadian males ($p = 1$), US males ($p = 2$), English and Welsh males ($p = 3$), Dutch males ($p = 4$) and West German males ($p = 5$), using the method of maximum likelihood with the following identifiability constraints: $\sum_{x=x_a}^{x_b} b_x^c = 1$,

Table 4.2: The estimates of the parameters in the transition equation of the ACF model.

$\vec{U}$		D		Q	
μ^c	-3.7903×10^{-1}	–	–	Q^c	1.0739×10^{-1}
$\mu^{(1)}$	-2.5276×10^{-3}	$\phi^{(1)}$	9.0759×10^{-1}	$Q^{(1)}$	1.5802×10^{-3}
$\mu^{(2)}$	-4.9883×10^{-3}	$\phi^{(2)}$	9.0310×10^{-1}	$Q^{(2)}$	1.4631×10^{-1}
$\mu^{(3)}$	-1.1591×10^{-1}	$\phi^{(3)}$	9.3293×10^{-1}	$Q^{(3)}$	5.9729×10^{-1}
$\mu^{(4)}$	6.1779×10^{-2}	$\phi^{(4)}$	9.4680×10^{-1}	$Q^{(4)}$	3.1499×10^{-1}
$\mu^{(5)}$	-3.3226×10^{-2}	$\phi^{(5)}$	9.1873×10^{-1}	$Q^{(5)}$	2.0034×10^{-1}

$\sum_{x=x_a}^{x_b} b_x^{(p)} = 1$ for $p = 1, \ldots, n_p$, $\sum_{t=t_a}^{t_b} k_t^c = 0$, and $\sum_{t=t_a}^{t_b} k_t^{(p)} = 0$ for $p = 1, \ldots, n_p$. The sample age range $[x_a, x_b]$ and calibration window $[t_a, t_b]$ used are $[60, 89]$ and $[1961, 2009]$, respectively. All required data are obtained from the Human Mortality Database (2015).

The estimates of the age-specific parameters in the observation equation and the hidden states over the calibration window are shown graphically in Figure 4.3. The estimates of the parameters in the transition equation are provided in Table 4.2.

In what follows, we set t_0 to the end of 2009, the last year of the calibration window. Also, following the assumptions made in Section 4.5.1, $P_L = 1$ and the possible values of P_H are 2, 3, 4 and 5.

4.5.3 Hedging Results I: Static Hedges

In this sub-section we consider static hedges that are established at time t_0 and are left unadjusted afterwards. Table 4.3 compares the resulting values of HE (calculated by simulation[9]) when population basis risk is absent and present. The value of HE is close to 80% in the ideal world where population basis risk is non-existent, but reduces to 30-60% in a more realistic situation when population basis risk exists (i.e., $P_H \neq P_L = 1$). It is interesting to note that the reduction in HE depends heavily on the choice of reference population: the reduction in HE is about 20 percentage points when the reference population is Dutch males ($P_H = 4$), but is close to 50 percentage points when the reference population is West German males ($P_H = 5$). This result suggests that it is important to choose the q-forward's reference population carefully.

Also shown in Table 4.3 are the values of $\widehat{HE}$, the analytical approximation of HE. For all five cases, the values of $\widehat{HE}$ are very close to the corresponding simulated values of

[9]The result is based on 10000 simulated paths.

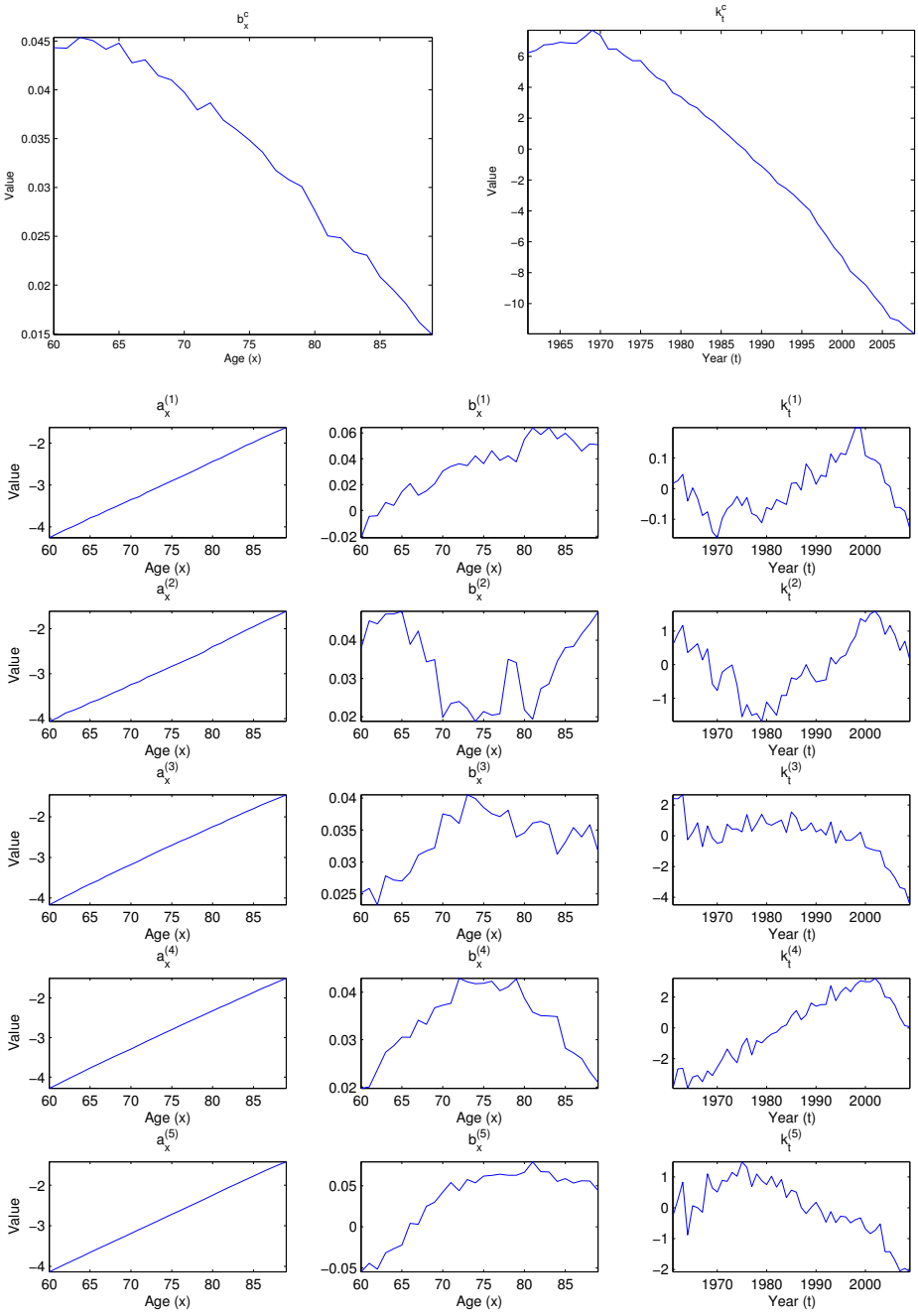


Figure 4.3: The estimates of the age-specific parameters $a_x^{(p)}$, b_x^c and $b_x^{(p)}$, $p = 1, \ldots, 5$, and the hidden states k_t^c and $k_t^{(p)}$, $p = 1, \ldots, 5$, in the ACF model.

Table 4.3: The values of HE (calculated by simulation) and $\widehat{HE}$ (the analytical approximation of HE) when population basis risk is absent and present.

	Basis risk is absent	Basis risk is present			
		$P_H = 2$	$P_H = 3$	$P_H = 4$	$P_H = 5$
HE (calculated by simulation)	0.7905	0.5393	0.3816	0.5794	0.3159
$\widehat{HE}$ (calculated analytically)	0.7902	0.5352	0.3860	0.5725	0.3155

Table 4.4: The component variances, $V_1(t_0), \ldots, V_5(t_0)$, when population basis risk is assumed to be absent.

	$V_1(t_0)$	$V_2(t_0)$	$V_3(t_0)$	$V_4(t_0)$	$V_5(t_0)$
An optimized hedge with 1 q-forward	0.0249	0.0197	−0.0393	0.0000	0.0000
No hedge	0.0249	0.0000	0.0000	0.0000	0.0000

HE, suggesting that the analytical approximation is reasonably accurate and may hence be used in practice to save computational effort.

We now move to studying the component risks: $V_1(t_0), \ldots, V_5(t_0)$. As discussed in Sections 4.3.2 and 4.3.3, they can be computed analytically. Let us first consider the situation when population basis risk is assumed to be absent (Table 4.4). In such a situation, all population-specific states are non-stochastic, so that $V_4(t_0) = V_5(t_0) = 0$ regardless of the q-forward's notional amount. When the liabilities are left unhedged, $V_2(t_0) = V_3(t_0) = 0$ while $V_1(t_0) = 0.0249 > 0$, so that the total risk is 0.0249. If an optimized hedge is in place, then $V_2(t_0)$ becomes positive but $V_3(t_0)$ becomes negative and larger than $V_2(t_0)$ in magnitude. In effect, the total risk, $V_1(t_0) + V_2(t_0) + V_3(t_0)$, is reduced to 0.0053 (which is significantly smaller than 0.0249).

When there is no population basis risk, the value of HE is 0.7905 and the optimized standardized notional amount $\hat{\mathcal{N}}_1^{(P_H)}(t_0)$ is 3.0537 no matter which reference population P_H is chosen. However, for the reasons provided in Section 4.4.2, the optimized (non-standardized) notional amounts do depend on P_H. In particular, we have $\hat{N}_1^{(2)}(t_0) = 395.0998$, $\hat{N}_1^{(3)}(t_0) = 476.3772$, $\hat{N}_1^{(4)}(t_0) = 489.7257$, $\hat{N}_1^{(5)}(t_0) = 401.9512$. These results can be visualized in Figure 4.4, which shows how $V_1(t_0) + V_2(t_0) + V_3(t_0)$ (i.e., the total longevity risk that the hedger is exposed to when there is no population basis risk) varies with $\mathcal{N}_1^{(P_H)}(t_0)$ and $N_1^{(P_H)}(t_0)$ for different choices of P_H.

Next, we consider a more realistic situation when population basis risk is present (Table 4.5). In this situation, $V_4(t_0)$ is positive and is invariant with both the notional amount of

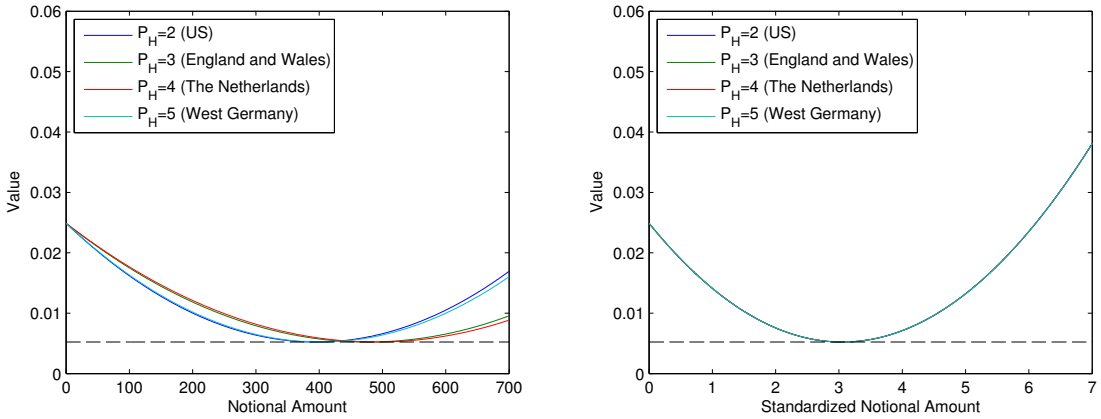


Figure 4.4: The relationship between $V_1(t_0)+V_2(t_0)+V_3(t_0)$ and $N_1^{(P_H)}(t_0)$ (the left panel), and the relationship between $V_1(t_0)+V_2(t_0)+V_3(t_0)$ and $\mathcal{N}_1^{(P_H)}(t_0)$ (the right panel); $P_H = 2, 3, 4, 5$.

the q-forward and the choice of the q-forward's reference population. The value of $V_4(t_0)$ (0.0001) reflects the portion of population basis risk arising from the hidden states that are associated with the hedger's own population P_L. On the other hand, $V_5(t_0)$ is zero when the liabilities are left unhedged but becomes positive as a q-forward contract is acquired. The value of $V_5(t_0)$ depends on P_H, reflecting the portion of population basis risk arising from the hidden states that are associated with the q-forward's reference population.

It can be observed from Table 4.5 that the choice of reference population has a huge impact on the values of $V_2(t_0)$, $V_3(t_0)$ and $V_5(t_0)$ and consequently the hedge effectiveness. This outcome may be understood by studying Figure 4.5, which shows the following three curves for each of the candidate reference populations:

(I) $V_1(t_0)+V_2(t_0)+V_3(t_0)$ against the standardized notional amount $\mathcal{N}_1^{(P_H)}(t_0)$;

(II) $V_4(t_0)+V_5(t_0)$ against the standardized notional amount $\mathcal{N}_1^{(P_H)}(t_0)$;

(III) $V_1(t_0)+V_2(t_0)+V_3(t_0)+V_4(t_0)+V_5(t_0)$ against the standardized notional amount $\mathcal{N}_1^{(P_H)}(t_0)$.

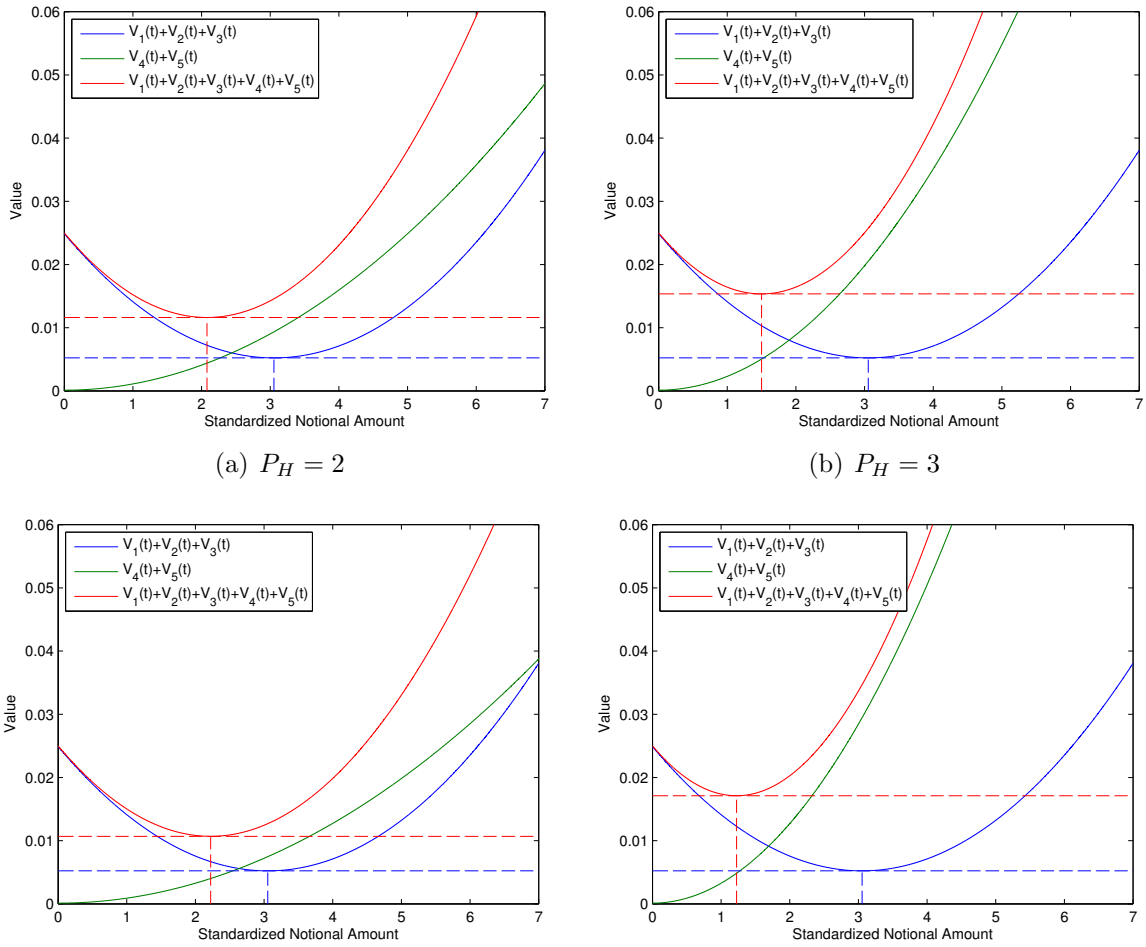
As previously explained, Curve (I) for all reference populations must be identical (see also the right panel of Figure 4.4). For different reference populations, the vertical intercepts of Curve (II) are the same (with a value of $V_4(t_0) = 0.0001$) but the curvatures of Curve (II) are not. Because Curve (III) is the superimposition of Curves (I) and (II), the

Table 4.5: The component variances, $V_1(t_0), \ldots, V_5(t_0)$, the optimized notional amount, $\hat{N}_1^{(P_H)}(t_0)$, the optimized standardized notional amount, $\hat{\mathcal{N}}_1^{(P_H)}(t_0)$, and the standardized basis risk profile, $BRP(x_1, T_1, P_H)$, for candidate reference populations $P_H = 2, 3, 4, 5$.

	$V_1(t_0)$	$V_2(t_0)$	$V_3(t_0)$	$V_4(t_0)$	$V_5(t_0)$	$\hat{N}_1^{(P_H)}(t_0)$	$\hat{\mathcal{N}}_1^{(P_H)}(t_0)$	$BRP(x_1, T_1, P_H)$
$P_H = 2$ (The United States)								
One q-forward	0.0249	0.0091	−0.0268	0.0001	0.0043	268.8725	2.0781	0.0010
Unhedged	0.0249	0.0000	0.0000	0.0001	0.0000			
$P_H = 3$ (England and Wales)								
One q-forward	0.0249	0.0047	−0.0193	0.0001	0.0049	233.7923	1.4986	0.0022
Unhedged	0.0249	0.0000	0.0000	0.0001	0.0000			
$P_H = 4$ (The Netherlands)								
One q-forward	0.0249	0.0104	−0.0286	0.0001	0.0039	356.4354	2.2225	0.0008
Unhedged	0.0249	0.0000	0.0000	0.0001	0.0000			
$P_H = 5$ (West Germany)								
One q-forward	0.0249	0.0032	−0.0158	0.0001	0.0047	161.2411	1.2250	0.0031
Unhedged	0.0249	0.0000	0.0000	0.0001	0.0000			

shape and position of Curve (III) depend on P_H. In particular, if the curvature of Curve (II) for a certain reference population is steeper, that is, the marginal increment in population basis risk per unit standardized notional amount acquired is higher, then the vertex of Curve (III) for the reference population would be positioned relatively higher and lefter. Therefore, for a reference population with a steeper Curve (II), the minimized variance is higher (which corresponds to a smaller value of HE) and the optimized standardized notional amount is smaller (which corresponds to smaller magnitudes of $V_2(t_0)$ and $V_3(t_0)$ as the magnitudes of these two components are positively related to $\mathcal{N}_1^{(P_H)}(t_0)$).

It is now clear that the relative degree of population basis risk depends entirely on the curvature of Curve (II). Figure 4.6 displays each candidate reference population's Curve (II). For each P_H, the order of the curvature of its Curve (II) is exactly the same as the order of its resulting population basis risk (measured in terms of the reduction in HE; see Table 4.3). Recall that the standardized basis risk profile $BRP(x_1, T_1, P_H)$ is defined as the curvature of Curve (II). Hence, instead of a graphical mean, one may use this quantity to choose the reference population that would lead to the smallest amount of population basis risk.



(a) $P_H = 2$ (b) $P_H = 3$

(c) $P_H = 4$ (d) $P_H = 5$

Figure 4.5: The curves of (I) $V_1(t_0) + V_2(t_0) + V_3(t_0)$, (II) $V_4(t_0) + V_5(t_0)$, and (III) $V_1(t_0) + V_2(t_0) + V_3(t_0) + V_4(t_0) + V_5(t_0)$ against the standardized notional amount $\mathcal{N}_1^{(P_H)}(t_0)$, $P_H = 2, 3, 4, 5$.

4.5.4 Hedging Results II: Dynamic Hedges

In this sub-section we consider dynamic hedges that are adjusted annually. In more detail, it is assumed that at time t_0, the hedger writes a $\hat{N}_1^{(P_H)}(t_0)$ notional q-forward with reference age x_1, time-to-maturity T_1 and reference population P_H. Then, at each future time t, where $t = t_0 + 1, \ldots, t_0 + T - 1$, the hedger closes out the q-forward written at $t - 1$ and

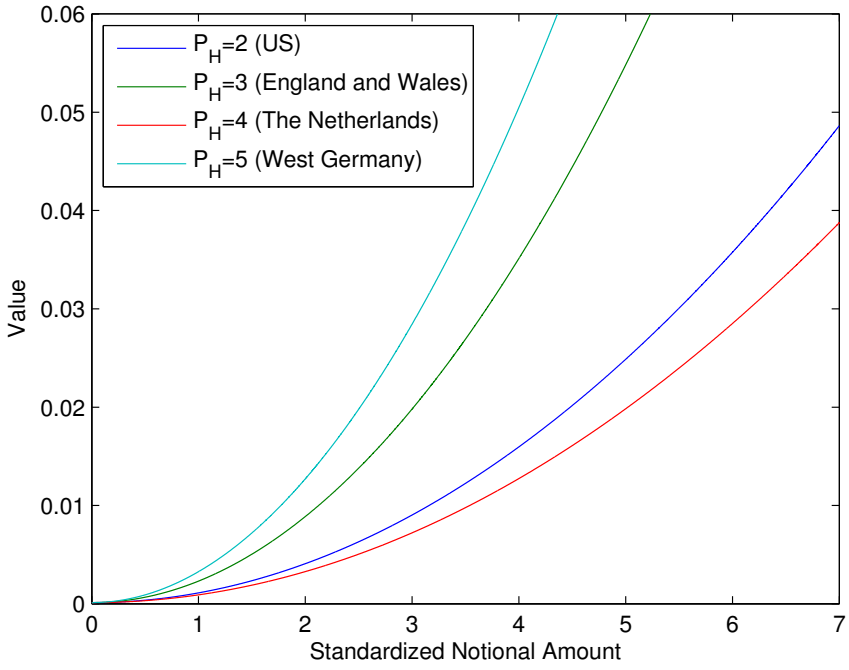


Figure 4.6: $V_4(t_0) + V_5(t_0)$ against the standardized notional amount $\mathcal{N}_1^{(P_H)}(t_0)$, $P_H = 2, 3, 4, 5$.

writes a $\hat{N}_1^{(P_H)}(t)$ notional freshly launched q-forward, which is characterized by the same parameters x_1, T_1 and P_H. Finally, at time $t = T$, all of the q-forwards in the hedge portfolio are closed out.

Table 4.6 compares the values of HE resulting from static and dynamic hedges. By dynamically adjusting a hedge, the value of HE can be increased by approximately 12-16 percentage points. The increase in HE due to dynamic adjustments does not seem to depend heavily on whether population basis risk exists.

Recall that the relative degree of population basis risk depends entirely on the standardized basis risk profile $BRP(x_1, T_1, P_H)$, which is defined exclusively by the q-forward parameters x_1, T_1 and P_H. Compared to the corresponding static hedge, the standardized basis risk profile applicable to a dynamic hedge is identical and remains constant (as we use q-forwards with parameters x_1, T_1 and P_H consistently over the hedge horizon). For this reason, the order of HE remains the same when we switch a hedge from static to dynamic.

To further analyze the hedges, in Figure 4.7 we display the simulated distributions of

Table 4.6: A comparison of the values of HE resulting from static hedges and the corresponding dynamic hedges.

	Basis risk is absent	Basis risk is present			
		$P_H = 2$	$P_H = 3$	$P_H = 4$	$P_H = 5$
Static hedge	0.7905	0.5393	0.3816	0.5794	0.3159
Dynamic hedge	0.9181	0.7001	0.5150	0.7068	0.4486

the time-t_0 values of the unexpected cash flows under the following conditions:

(i) when the liabilities are left unhedged, i.e., $L(t_0) - \mathrm{E}(L(t_0)$

(ii) when the liabilities are statically hedged, i.e., $L(t_0) - \mathrm{E}(L(t_0)) - N_1^{(P_H)}(t_0)H_1^{(P_H)}(t_0)$;

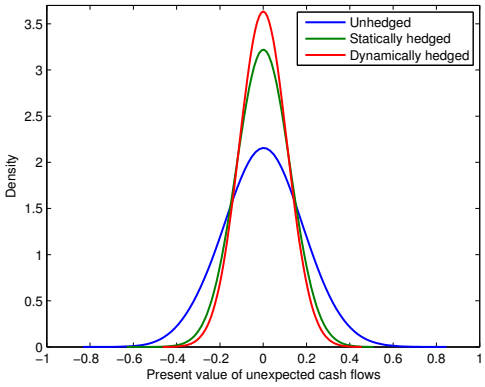
(iii) when the liabilities are dynamically hedged, i.e., $L(t_0) - \mathrm{E}(L(t_0)) - \sum_{t>t_0} PCF(t)|\mathcal{F}_{t_0}$.

All of the three distributions are centred at zero, as the expectations of $L(t_0) - \mathrm{E}(L(t_0))$, $H_1^{(P_H)}(t_0)$ and $\sum_{t>t_0} PCF(t)|\mathcal{F}_{t_0}$ are zero. No matter if population basis risk exists, Distribution (i) is the most dispersed while Distribution (iii) is the least. This observation indicates that a static hedge reduces the variability of the unexpected cash flows and that the corresponding dynamic hedge reduces the variability even more. Of course, the reductions in the variability of the unexpected cash flows are the most prominent when population basis risk is assumed to be absent. When population basis risk is present, the reductions in the variability of the unexpected cash flows depend quite significantly on P_H.

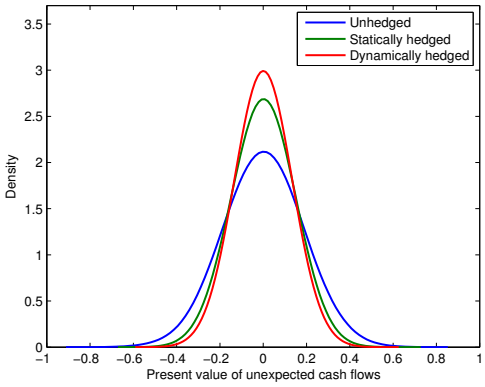
4.6 Concluding Remarks

It is believed that hedgers' concern about population basis risk is a major obstacle to the development of the market for index-based longevity risk transfers. In 2013, the LLMA established the Longevity Basis Risk Working Group, whose goal is to develop readily applicable methods for quantifying population basis risk. This chapter addresses the needs of the market by introducing a method for measuring population basis risk efficiently and a framework for optimizing index-based longevity hedges when the index used is associated with a population that is different from the hedger's own population.

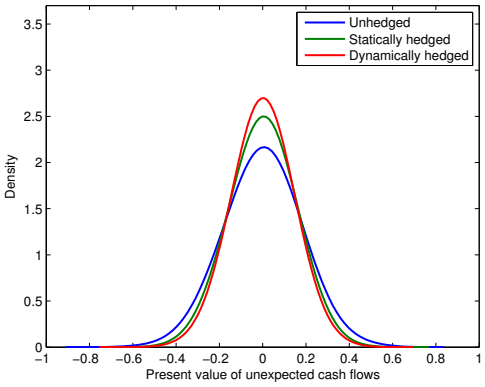
The first contribution of this chapter is an extension of the generalized state-space hedging method proposed in Chapter 3. The extension allows us to calibrate a longevity hedge



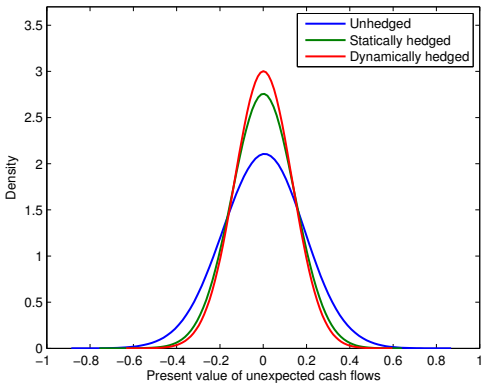
No population basis risk



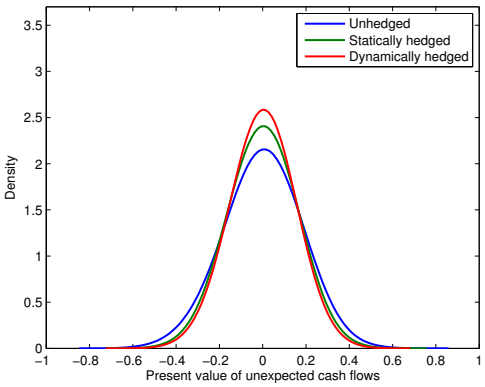
$P_H = 2$



$P_H = 3$



$P_H = 4$



$P_H = 5$

Figure 4.7: The simulated distributions of the time-t_0 values of the unexpected cash flows when the liabilities are unhedged, statically hedged and dynamically hedged.

that is composed of standardized hedging instruments, taken into account the difference in mortality improvements between the populations associated with the hedging instruments and the hedger's portfolio. The extension may be regarded as a general framework, because it is applicable to all coherent multi-population stochastic mortality models that can be written in state-space representations. We have demonstrated empirically that when calibrated by the proposed method, a static longevity hedge with one q-forward can yield an effectiveness of 30-60%, depending on the choice of the reference population. The effectiveness would become 12 to 16 percentage points higher if the hedge is adjusted annually over the hedging horizon.

The second contribution is a method to analytically approximate the variances of $L(t)$ (the time-t value of unhedged position) and $L(t) - \sum_{j=1}^{m} N_j^{(P_H)}(t) H_j^{(P_H)}(t)$ (the time-t value of the hedged position). This approximation method enables us to calculate the hedge effectiveness,

$$HE = 1 - \frac{\text{Var}\left(L(t_0) - \sum_{j=1}^{m} N_j^{(P_H)}(t_0) H_j^{(P_H)}(t_0)\right)}{\text{Var}(L(t_0))},$$

without using simulation. Using this approximation method, we can derive an empirical relationship between HE and the composition of the hedge portfolio $(N_1^{(P_H)}(t), \ldots, N_m^{(P_H)}(t))$ with only minimal computational effort. Furthermore, by applying this approximation method to the full multi-population mortality model and the restricted model with $Q^{(1)} = Q^{(2)} = \ldots = Q^{(n_p)} = 0$, we can readily gauge to what extend population basis risk erodes hedge effectiveness under different hedge portfolio compositions.

The third contribution is an analysis of longevity risk through a variance decomposition. We have shown that the (approximated) variance of the time-t values of a hedged annuity portfolio can be expressed as the sum of five components, $V_1(t)+V_2(t)+V_3(t)+V_4(t)+V_5(t)$. We have argued that $V_4(t)$ and $V_5(t)$ collectively represent an explicit measure of the population basis risk involved in the longevity hedge. In particular, $V_4(t)$ represents the portion of population basis risk that exists no matter how the hedge is composed, whereas $V_5(t)$ represents the portion of population basis risk that depends on the composition of the hedge portfolio. We have also shown that $V_5(t)$ increases quadratically with the standardized notional amounts $\mathcal{N}_1^{(P_H)}(t), \ldots, \mathcal{N}_m^{(P_H)}(t)$ of the q-forwards used, and that the curvature depends on the parameters of the q-forwards. Using these properties, we define a metric called standardized basis risk profile, which allows us to assess the relative levels of population basis risk that q-forwards with different reference populations, reference ages and times-to-maturity may lead to.

We conclude with the caveats of our work and suggestions for future research. The

hedge effectiveness reported in this chapter does not take model and parameter risks into account. If the true underlying model is not identical to the model used in deriving the general state-space hedging strategy, the hedge effectiveness may be reduced. To estimate the possible reduction, one may recalculate hedge effectiveness by simulating mortality scenarios generated by alternative multi-population mortality models. Similarly, the impact of parameter risk can be estimated by recomputing hedge effectiveness using mortality scenarios that are generated by bootstrapping or Bayesian methods (Brouhns et al., 2005; Cairns et al., 2011b; Yang et al., 2015).

It should also be noted that small sample risk (a.k.a. Poisson risk and sampling risk) is not incorporated into the results of this chapter. One may estimate the impact of small sample risk by simulation approaches, in which realizations of death counts are simulated from a certain counting distribution such as Poisson and binomial (Cairns et al., 2014; Li and Hardy, 2011). It may also be possible to estimate the impact of small sample risk by analytical means. In more detail, when small sample risk is considered, the variance of $L(t)$ would consist additionally a component representing the variance of the time-t values of the future liabilities given known values of $q^{(P_L)}_{x_0+t-t_0}, q^{(P_L)}_{x_0+t+1-t_0}, \ldots$. If this component can be analytically computed or approximated, then we can obtain an analytical approximation of HE that incorporates small sample risk. The derivation of such an analytical expression is left for future research.

Chapter 5

An Efficient Method for Hedging Period and Cohort Effects in Longevity Risk

5.1 Introduction

Solutions for hedging pure longevity risk can be broadly divided into two categories: customized and index-based. A customized longevity hedge is based on the actual mortality experience of the individuals associated with the liability being hedged, whereas an index-based longevity hedge is linked to a broad-based mortality index which reflects the actual mortality experience of a larger pool of individuals, such as a national population. Each of these two types of longevity hedge has its pros and cons.

A customized hedge eliminates all longevity risk, but it is more costly and difficult to unwind. It is also less attractive to capital markets investors, because it lacks liquidity and transparency. Typically, a customized hedge is structured as a 'cash flow hedge', in which net payments are made period by period to immunize the longevity risk associated with every single cash flow of the hedger. An example of a customized cash flow hedge is the longevity swap that was executed between J.P. Morgan and Canada Life in 2009. Under this 40-year longevity swap, Canada Life receives from J.P. Morgan the actual payments it must pay to its annuitants and, in return, makes a series of fixed payments to J.P. Morgan (Blake et al., 2013).

In contrast, an index-based hedge is not perfect, leaving behind residual risks such as small sample risk and population basis risk. The problems arising from these residual risks

have been analyzed by researchers including Cairns et al. (2014), Coughlan et al. (2011), Haberman et al. (2014), Li and Hardy (2011) and Ngai and Sherris (2011). Nevertheless, an index-based hedge is potentially more liquid and consequently cheaper due to a lower illiquidity premium. If the hedge is adjusted periodically, then the hedging instruments used can be shorter-dated, better meeting investors' preference and reducing the hedger's exposure to counterparty default risk (see, e.g., Cairns, 2011; Zhou and Li, 2014). Furthermore, because the value of a pension or annuity liability depends on the level of the relevant mortality index at the time of valuation, index-based hedges can be structured as a 'value hedge', which immunizes the risk associated with the liability's value rather than the liability's cash flows. An example of an index-based value hedge is the q-forward contract that Lucida executed with J.P. Morgan to lock in the value of its annuity liability at the end of the hedging horizon (Blake et al., 2013).

As Coughlan et al. (2013) pointed out, index-based value hedges are very well suited for de-risking deferred liabilities, which involve no cash flow during the period of deferral. They permit pension plan sponsors with a large number of active members (who are still years from receiving their pension benefits) to offload their longevity risk exposures. As an index-based value hedge is quite flexible, it can be adjusted from time to time according to, for example, the changes in the composition of plan sponsor's workforce and the options (e.g., a lump sum vs. lifetime payments) chosen by the plan members. Index-based value hedges also provide a way for insurers selling advanced-life delayed annuities (i.e., deeply deferred life annuities) to reduce their longevity risk exposures and hence the associated solvency capital.[1] The world's first index-based value hedge for deferred members of a pension plan was the q-forward deal executed between the Pall (UK) Pension Fund and J.P. Morgan in 2011 (Blake et al., 2014). The hedge was calibrated to lock in the value of the deferred pension liability over a 10-year horizon.

Although the suitability for deferred pension and annuity liabilities is an important feature of index-based value hedges, this feature has not been extensively studied in the literature. From a technical viewpoint, deferred liabilities often involve cohorts that are not covered by the data sample to which the underlying mortality model is calibrated. To illustrate, let us suppose that the underlying mortality model is calibrated to data over ages 50 to 89 (a typical age range to which mortality models for pension valuation are fitted) and a sample period that includes the current year. The data sample does not cover the cohorts who are now age 49 or younger, which means when valuing the deferred

[1] Advanced-life delayed annuities (ALDAs) are also known as longevity insurance, deferred income annuities (DIAs) and qualified longevity annuity contracts (QLACs). In the US, ALDAs are available from insurers including Guardian, New York Life, Mass Mutual, MetLife and Pacific Life. We refer interested readers to Gong and Webb (2010), Horneff et al. (2010) and Milevsky (2005) for further information concerning ALDAs.

liabilities for these cohorts, the cohort effects in the model must be projected. In other words, the deferred liabilities are subject to the stochastic uncertainty surrounding not only period (time-related) effects but also cohort (year-of-birth-related) effects. However, cohort effects are not taken into account in most of the existing methods for calibrating index-based hedges, including the method discussed in Chapter 3 and those proposed by Blake et al. (2006), Cairns (2011), Dahl (2004), Dahl and Møller (2006), Dahl et al. (2008), Zhou and Li (2014) and Luciano et al. (2012).

So far as we aware, Li and Luo (2011), Cairns et al. (2014) and Cairns (2013) have considered cohort effects in the context of hedging. Li and Luo (2011) incorporated the potential dependence between different cohorts into a static longevity hedge, but their set-up is a cash flow hedge that is not particularly suitable for a deferred liability. Cairns et al. (2014) and Cairns (2013) studied a value hedge for a deferred annuity that is payable to a single cohort of individuals; however, in their set-up, the cohort in question is covered by the data sample used in calibrating the simulation model, which means that the annuity liability is not subject to any cohort effect uncertainty. Therefore, their results do not indicate how cohort effect uncertainty may be mitigated if it is involved in liability being hedged. Another shortcoming is that their calculations are based on the assumption that all possible (simulated) outcomes are known a priori. This approach provides the best achievable hedge effectiveness given the range of simulated outcomes, but a simulation-based calibration could be computationally intensive, especially when multiple hedging instruments are included in the hedge portfolio.

This chapter complements the literature by contributing a method for hedging the uncertainty surrounding both period and cohort effects. Using the proposed method, one can create a value hedge for a deferred annuity liability which involves cohort effects that are not yet realized as of the time when the hedge is established. The hedging instruments used are q-forwards, which may be linked to cohorts that are different from those that are associated with the annuity liability. At the user's discretion, the hedge can be executed as a static hedge (which remains unchanged over the hedging horizon) or a dynamic hedge (which is adjusted periodically over the hedging horizon). The proposed method is developed from the stochastic properties of the innovations in the assumed processes for the underlying period and cohort effects. It yields hedge ratios that can be expressed analytically in terms of (i) the variances and covariances of the innovations and (ii) the partial derivatives of the values of the hedge portfolio and the annuity liability with respect to the relevant innovations. As no simulation is required in the calculating the hedge ratios, the execution of our proposed hedging method requires minimal computational effort.

Unlike the hedging strategies proposed by Blake et al. (2006), Cairns (2011), Li and Luo (2011) and Zhou and Li (2014), the hedge ratios under our method are not obtained

simply by matching the sensitivities of the liability and the hedge portfolio to changes in the underlying mortality. Instead, our method has a closer resemblance to the generalized state-space hedging method proposed in Chapter 3, under which the optimal hedge ratios are calculated on the basis of a specific risk measure.[2] Compared to the sensitivity matching approaches, this approach to deriving hedge ratios is more suitable for hedgers who have a definite hedging objective. The risk measures we consider include variance and Value-at-Risk (relative to the unhedged expected value). The former is commonly used for the purpose of evaluating the effectiveness of longevity hedges (see Cairns, 2011, 2013; Cairns et al., 2014; Coughlan et al., 2011; Li and Hardy, 2011), while the latter is more in line with the calculation of solvency risk capital under Solvency II (see Plat, 2011; Olivieri and Pitacco, 2009). We illustrate the proposed hedging method using real mortality data from a national population. The empirical work demonstrates the benefit of factoring cohort effects into a longevity hedge under different circumstances.

Another objective of this chapter is to address the computational burden involved when assessing the performance of a value longevity hedge. Generally speaking, to evaluate the performance of a value longevity hedge, nested simulations (i.e., simulations on simulations) are required. For instance, to evaluate a value longevity hedge with a hedging horizon of τ years, we first simulate, say M_1, mortality paths over the first τ years since the hedge is established. Then, for each of these M_1 mortality paths, we need to calculate the value of the hedged position at the end of the hedging horizon, which demands another, say M_2, sample mortality paths (for years $\tau + 1$ and beyond). In total, $M_1 \times M_2$ sample paths are required, and therefore the computation burden is significant as both M_1 and M_2 are large numbers. It has been proposed to use approximation methods to eliminate the need for the M_2 sample paths (Cairns, 2011; Cairns et al., 2014; Denuit and Dhaene, 2007; Denuit et al., 2010; Zhou and Li, 2014).

In this chapter, we attack the problem of nested simulations from a different angle. Instead of using an approximation, we utilize the statistical properties of the innovations from the underlying period and cohort effect processes. We recognize that the value of the hedge portfolio at time τ (measured from the time at which the hedge is established) can be reformulated in such a way that the only random components are the innovations beyond time τ. On the basis of this fact, we evaluate hedge effectiveness by simulating sample paths of innovations rather than sample paths of period and cohort effects. Because

[2]The generalized state-space hedging method requires the user to formulate the underlying stochastic mortality model in a state-space representation, comprised of an observation equation that relates the observed mortality to the latent state factors and a first order state process that governs the evolution of the latent state factors over time. It is not straightforward to express a stochastic mortality model in a state-space representation if it contains a cohort effect term. Therefore, the generalized state-space hedging method cannot be applied readily to the problem considered in this chapter.

all innovations are serially uncorrelated by definition, the M_2 sample paths for valuation purposes at time τ do not depend on the M_1 sample paths in the first stage, thereby sparing us from the need of nested simulations. As the proposed method does not depend on any approximation, it is not subject to the problem of approximation accuracy and the resulting hedge effectiveness measure can better reflect the 'reality' implied by the simulated mortality scenarios.

The remainder of this chapter is organized as follows. Section 5.2 presents the stochastic mortality model that is assumed throughout this chapter. Section 5.3 describes the liability being hedged and the hedging instruments. Section 5.4 introduces the proposed hedging method and details how the hedge ratios under the proposed method can be calculated. Section 5.5 explains our proposed method for evaluating value hedges. Section 5.6 presents the baseline empirical results and performs several sensitivity tests. Finally, Section 5.7 concludes the chapter with some suggestions for future research.

5.2 The Assumed Model

5.2.1 Specification

The model we use throughout the chapter is Model M7, a generalized version of the original Cairns-Blake-Dowd model (Cairns et al., 2006). It has been shown to perform satisfactorily when applied to several national populations including the United States and England and Wales (Cairns et al., 2009, 2011a; Dowd et al., 2010a,b).

Let $q_{x,t}$ be the probability that an individual dies between time $t-1$ and t (i.e., during year t), given that he/she has survived to age x at time $t-1$. Model M7 captures $q_{x,t}$ as follows:

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{(4)}, \tag{5.1}$$

where $\bar{x}$ is the average of the age range $[x_a, x_b]$ to which the model is calibrated and $\hat{\sigma}_x^2$ is the mean of $(x-\bar{x})^2$ over the same age range. In the model, $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, and $\kappa_t^{(3)}$ are the period (time-related) effects, governing the random evolution of $q_{x,t}$ over time t. On the other hand, $\gamma_{t-x}^{(4)}$ is the cohort (year-of-birth) effect, which determines the random evolution of $q_{x,t}$ over the year of birth $(t-x)$ dimension.

Following Cairns et al. (2009), we estimate equation (5.1) by the method of Poisson

maximum likelihood with the following identifiability constraints:

$$\sum_{x,t} \gamma^{(4)}_{t-x} = 0, \quad \sum_{x,t} (t-x)\gamma^{(4)}_{t-x} = 0, \quad \text{and} \quad \sum_{x,t} (t-x)^2 \gamma^{(4)}_{t-x} = 0,$$

where the summations are taken over the entire age range $[x_a, x_b]$ and sample period $[t_a, t_b]$ to which the model is calibrated. These constraints are sufficient to stipulate parameter uniqueness. They also ensure that the estimated values of $\gamma^{(4)}_{t-x}$ will fluctuate around zero and will have no discernible linear or quadratic trend.

Having estimated equation (5.1) to historical death and exposure counts, the period effects are modeled jointly by a trivariate random walk with drift:

$$\begin{cases} \kappa^{(1)}_t &= \mu^{(1)} + \kappa^{(1)}_{t-1} + \eta^{(1)}_t \\ \kappa^{(2)}_t &= \mu^{(2)} + \kappa^{(2)}_{t-1} + \eta^{(2)}_t \\ \kappa^{(3)}_t &= \mu^{(3)} + \kappa^{(3)}_{t-1} + \eta^{(3)}_t \end{cases}, \tag{5.2}$$

where $\mu^{(1)}, \mu^{(2)}, \mu^{(3)}$ are the drift terms, and $\eta^{(1)}_t$, $\eta^{(2)}_t$ and $\eta^{(3)}_t$ are the random innovations. The vector of the innovations $(\eta^{(1)}_t, \eta^{(2)}_t, \eta^{(3)}_t)'$ possesses no serial correlation and follows a trivariate normal distribution with a zero mean vector and a constant (positive definite) covariance matrix Q. Because of the identifiability constraints used, the cohort effect $\gamma^{(4)}_{t-x}$ should have no long-term trend and can thus be modeled by a mean-reverting ARMA(R,M) process:

$$\gamma^{(4)}_{t-x} = \mu^{(4)} + \sum_{i=1}^{R} \phi^{(4)}_i \gamma^{(4)}_{t-x-i} + \eta^{(4)}_{t-x} + \sum_{i=1}^{M} \theta^{(4)}_i \eta^{(4)}_{t-x-i}, \tag{5.3}$$

where $\mu^{(4)}$ is the constant term, $\phi^{(4)}_1, \ldots, \phi^{(4)}_R$ are the AR coefficients, and $\theta^{(4)}_1, \ldots, \theta^{(4)}_M$ are the MA coefficients. The innovation $\eta^{(4)}_{t-x}$ possesses no serial correlation and follows a normal distribution with a zero mean and a constant variance of $(\sigma^{(4)})^2$.

5.2.2 Estimation

The empirical work in this chapter is based on the historical data from the male population of England and Wales (EW). The data are obtained from the Human Mortality Database (2015). We estimate equation (5.1) to the death and exposure counts for EW males over

the age range of $[x_a, x_b] = [50, 89]$ and the sample period of $[t_a, t_b] = [1941, 2011]$.[3]

The maximum likelihood estimates of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and $\gamma_{t-x}^{(4)}$ are shown graphically in Figure 5.1. The downward trend in $\kappa_t^{(1)}$ reflects a steady reduction in mortality at all ages over the sample period. The patterns of $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ respectively indicate how the gradient and curvature of the relationship between $\ln(q_{x,t}/(1-q_{x,t}))$ and x have changed since 1941. As expected, the estimated values of $\gamma_{t-x}^{(4)}$ fluctuate around zero with no observable trend. The spike at $t - x = 1918$ could attributed to the Spanish flu epidemic, which might have affected people who were born in 1918 (see, e.g., Holmes, 2004).

We have the estimated (realized) values of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ for $t = t_a = 1941$ to $t = t_b = 2011$, but the values of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ for $t > t_b$ are random, following the trivariate random walk specified in equation (5.2). For $t > t_b$, all values of $q_{x,t}$ are subject to the uncertainty surrounding the associated unrealized period effects. However, not all values of $q_{x,t}$ beyond the end point t_b of the sample period are subject to cohort effect uncertainty. In particular, cohort effect uncertainty only matters to those with $t - x > t_b - x_a = 2011 - 50 = 1961$, because the data sample covers years of birth up to $t_b - x_a = 1961$ and consequently the cohort effects for $t - x \leq 1961$ are already realized.[4] It follows that cohort effect uncertainty is irrelevant to liabilities that are payable to individuals born in or before 1961, but plays a role in deferred pension liabilities for workforces containing individuals born after 1961.

Table 5.1 reports the estimates of the parameters in the processes for the period and cohort effects. Parameters Q, $\theta_1^{(4)}$, $\phi_1^{(4)}$ and $\sigma^{(4)}$ are involved in the hedging strategy presented in Section 5.4. Also, the empirical fact that $Q(1,1)$ and $(\sigma^{(4)})^2$ are significantly larger than $Q(2,2)$ and $Q(2,3)$ is used to explain some interesting properties of the hedging results in Section 5.6.[5]

[3]We use the age range of [50, 89] for several reasons. First, mortality improvement dynamics over different age ranges are very different; given the purpose of the model is to value pension liabilities, it is more appropriate not to include younger ages. Second, the age functions, $x - \bar{x}$ and $(x - \bar{x})^2 - \hat{\sigma}_x^2$, in Model M7 do not capture the 'accident hump' at younger ages. Third, the data beyond age 90 are subject to reliability issues. In particular, the population counts for EW males beyond age 89 are extrapolated rather than raw values. As a fact, the age range of $[50, 89]$ is used in other studies including the work of Cairns et al. (2009).

[4]For instance, $q_{50,2021}$ is subject to the uncertainty surrounding $\gamma_{1971}^{(4)}$, but $q_{80,2021}$ is free of cohort effect uncertainty as $\gamma_{1941}^{(4)}$ is already known from the data sample.

[5]We use $Q(i,j)$ to denote the (i,j)th element of Q.

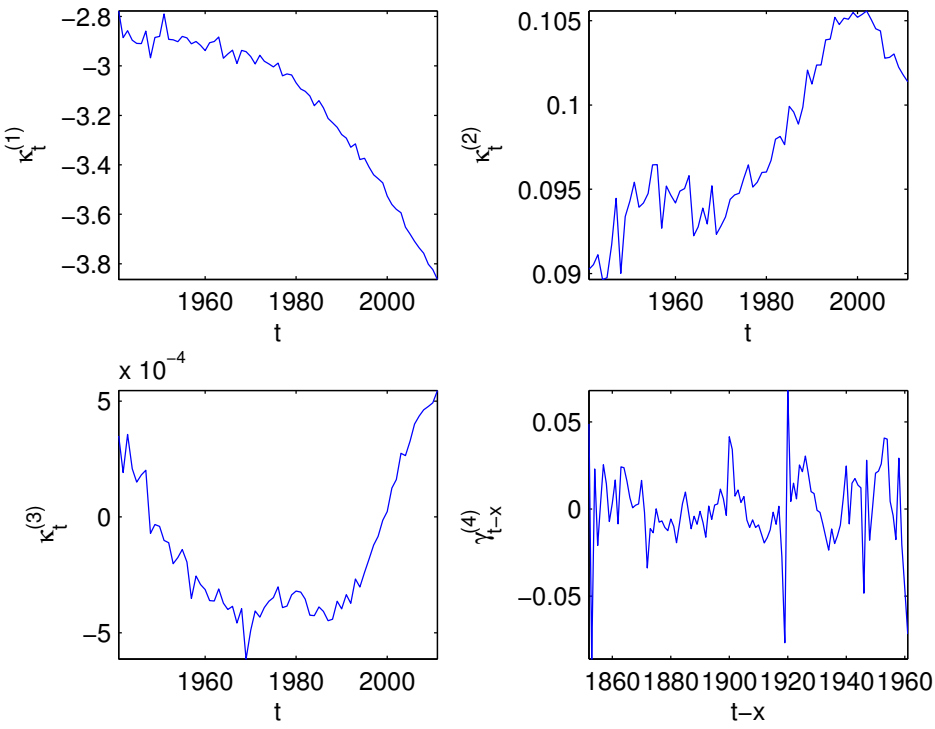


Figure 5.1: The Poisson maximum likelihood estimates of $\kappa_t^{(1)}$, $\kappa_t^{(2)}$, $\kappa_t^{(3)}$ and $\gamma_{t-x}^{(4)}$ in Model M7, $t = 1941, \ldots, 2011$ and $x = 50, \ldots, 89$.

5.2.3 Significance of Cohort Effects

To understand the importance of considering cohort effects, we also estimate the original Cairns-Blake-Dowd model (Model M5), which does not take cohort effects into account. Figure 5.2 compares the actual historical values of $\ln(q_{x,t}/(1-q_{x,t}))$ with the corresponding fitted values that are calculated from Models M5 and M7. It can be observed that the surface representing the actual historical values is not flat, with several humps and bumps along the diagonal dimension. The humps and bumps, which indicate that mortality varies across different years of birth, can be captured in the fitted values calculated from Model M7 (the model used in this chapter) but not in those calculated from Model M5 (the more restrictive model that ignores cohort effects).

The need for considering cohort effects can also be understood by considering the

Table 5.1: The estimates of the parameters in the trivariate random walk for $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ and the ARMA(1,1) process for $\gamma_{t-x}^{(4)}$. The ARMA order $(R = 1, M = 1)$ is identified using the Box-Jenkins method (Box and Jenkins, 1976).

The trivariate random walk for $\kappa_t^{(1)}$, $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$			
$\mu^{(1)}$	-0.0155		
$\mu^{(2)}$	1.5908×10^{-4}		
$\mu^{(3)}$	2.8048×10^{-6}		
Q	0.0013	3.4174×10^{-5}	9.3595×10^{-7}
	3.4174×10^{-5}	1.9104×10^{-6}	6.2427×10^{-8}
	9.3595×10^{-7}	6.2427×10^{-8}	5.8429×10^{-9}

The ARMA(1,1) process for $\gamma_{t-x}^{(4)}$							
$\mu^{(4)}$	-2.6647×10^{-4}	$\theta_1^{(4)}$	-0.4088	$\phi_1^{(4)}$	0.5612	$(\sigma^{(4)})^2$	4.9418×10^{-4}

standardized residuals from each mortality model:

$$Z_{x,t} = \frac{D_{x,t} - \hat{D}_{x,t}}{\sqrt{\hat{D}_{x,t}}},$$

where $D_{x,t}$ is the actual death count at age x and in year t, and $\hat{D}_{x,t}$ is the corresponding death count estimated from the model. If the model is adequate, then the pattern of $Z_{x,t}$ should be random and have little clustering. In Figure 5.3 we compare the heat maps of $Z_{x,t}$ from Models M5 and M7. For Model M7, the pattern of $Z_{x,t}$ looks reasonably random, indicating that the model gives an adequate fit to the historical data. However, for Model M5, the pattern of $Z_{x,t}$ contains some obvious diagonal bands, indicating a lack of fit which arises from features that are year-of-birth specific.

Figures 5.2 and 5.3 reminiscent the results produced by the Continuous Mortality Investigation Bureau (2002). They point to the conclusion that in the United Kingdom, where most of the longevity risk transfers in the world take place, cohort effects are significant and should be modeled explicitly. In the following sections, we explain how cohort effect uncertainty can be mitigated by an index-based longevity hedge, and discuss the importance of cohort effects from the perspective of hedge effectiveness.

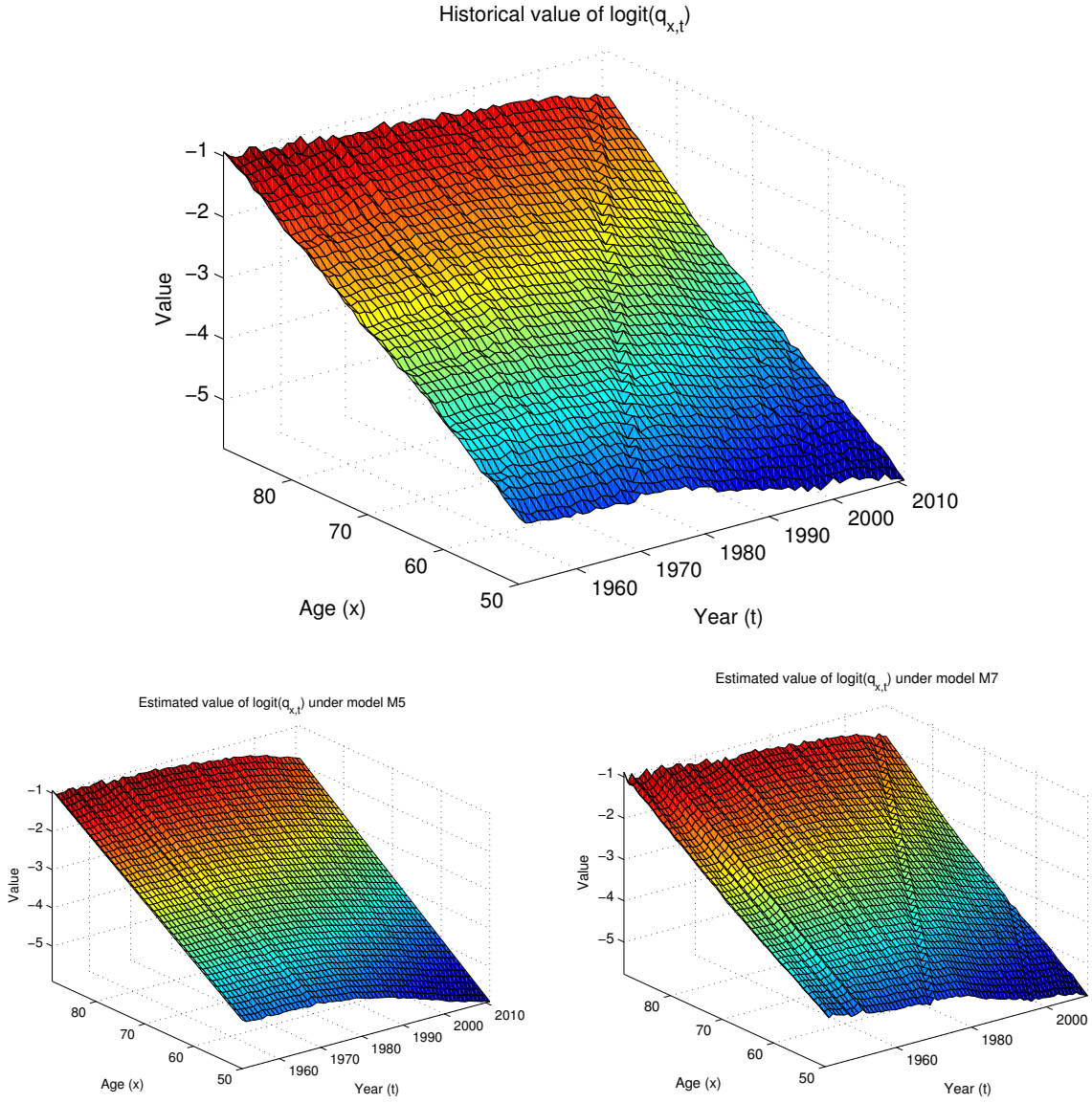


Figure 5.2: Historical values of $\ln(q_{x,t}/(1-q_{x,t}))$ and the corresponding fitted values that are calculated from Models M5 and M7.

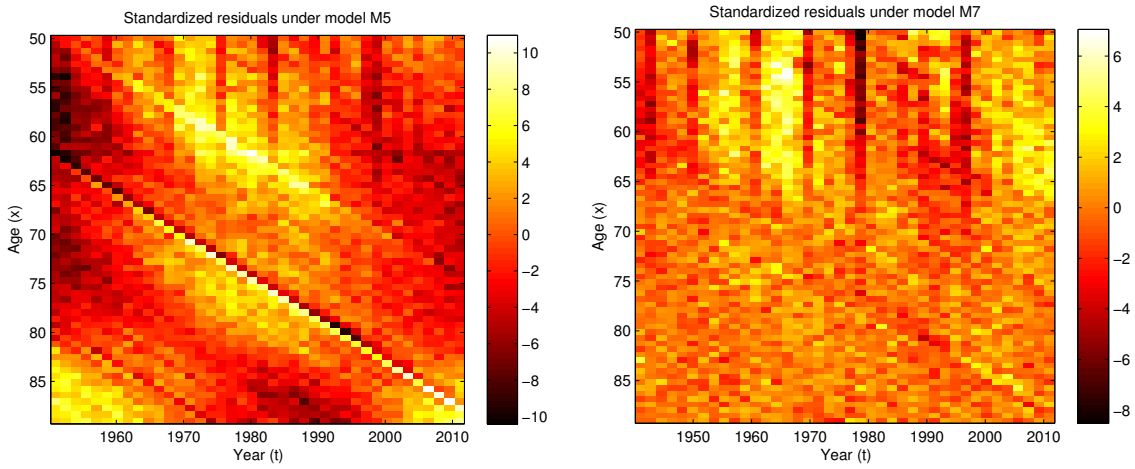


Figure 5.3: Heat maps of the standardized residuals $Z_{x,t}$ calculated from Models M5 and M7.

5.3 The Set-up

5.3.1 The Liability Being Hedged

The liability being hedged is a portfolio of T-year deferred life annuity immediate contracts, which are sold to individuals age $x_0 - T$ at time t_b (the end point of the data sample period). Each contract makes no payment during the first T years from time t_b; however, when the annuitant reaches age x_0 at time $t_b + T$, the contract pays \$1 at the end of each year (starting in year $t_b + T + 1$) as long as the annuitant is alive. We let L be the sum of the discounted cash flows, measured at time t_b, per contract. Ignoring sampling risk, we have

$$L = \sum_{u=1}^{\omega - x} e^{-r(T+u)} \prod_{s=1}^{T+u} (1 - q_{x_0 - T + s - 1, t_b + s}), \tag{5.4}$$

where r is the risk-adjusted discount rate and ω is the highest attainable age. It can be seen that L contains the underlying death probabilities $q_{x_0-T+s-1,t_b+s}$, for $s = 1, \ldots, T + \omega - x_0$. Under our model assumption, this in turn means that L is a function of a sequence of period effects, $\{\kappa^{(i)}_{t_b+s};\ i = 1, 2, 3,\ s = 1, \ldots, T + \omega - x_0\}$, and a cohort effect $\gamma^{(4)}_{t_b-(x_0-T)+1}$.

We let $L_t = L|\mathcal{F}_t$, where $t \geq t_b$ and $\mathcal{F}_t$ represents the information up to and including time t. The period and/or cohort effect uncertainty affecting L_t depends on t and the parameters of the annuity liability.

For $t_b \leq t < t_b + T + \omega - x_0$, L_t is subject to the randomness of the following period effects: $\kappa^{(i)}_{t+1}, \ldots, \kappa^{(i)}_{t_b+T+\omega-x_0}$, where $i = 1, 2, 3$. For $t > t_b + T + \omega - x_0$ (i.e., when all annuity payments have already been made), all period effects contained in L are realized and therefore L_t is no longer subject to any period effect uncertainty.

Recall that x_a is the lower limit of the sample age range. At $t \geq t_b$, we have the realizations of the cohort effects up to and including year-of-birth $t - t_a$. As such, whether L_t is subject to cohort effect uncertainty depends on the value of $t_b - (x_0 - T) + 1$ relative to that of $t - x_a$. If $t_b - (x_0 - T) + 1 > t - x_a$, then L_t is subject to the uncertainty associated with the unrealized value of $\gamma^{(4)}_{t_b-(x_0-T)+1}$. If $t_b - (x_0 - T) + 1 \leq t - x_a$, then $\gamma^{(4)}_{t_b-(x_0-T)+1}$ is already realized and thus L_t is free from any cohort effect uncertainty.

Suppose that the hedge is built at time t, where $t \geq t_b$. The longevity hedge we develop is a value hedge, mitigating the uncertainty associated with

$$V_{L_t}(t+\tau) := \mathrm{E}\left(L_t | \mathcal{F}_{t+\tau}\right), \tag{5.5}$$

the value of the liability $\tau > 0$ years from the time when the hedge is established.[6] We assume $\tau < T + \omega - x_0$, as it is not reasonable to have a hedging horizon that ends after all liability cash flows are paid.

As of time t, $V_{L_t}(t+\tau)$ is subject to uncertainty surrounding the random sample path of the period effects from time t to $t+\tau$, that is, $\kappa^{(i)}_{t+1}, \ldots, \kappa^{(i)}_{t+\tau}$ for $i = 1, 2, 3$. When $t_b - (x_0 - T) + 1 > t - x_a$, it is also subject to at least part of the sample path of the cohort effects over years of birth $t + 1 - x_a$ to $t + \tau - x_a$, that is, $\gamma^{(4)}_{t+1-x_a}, \ldots, \gamma^{(4)}_{t+\tau-x_a}$.[7]

5.3.2 The Hedging Instruments

The longevity hedge we develop consists of $m \geq 1$ q-forwards. Suppose that the q-forwards in the portfolio are launched at time t, where $t \geq t_b$. For the jth q-forward, the discounted net cash flow, measured at time t and from the perspective of the fixed-rate receiver, is

$$H(j,t) = e^{-rT_j}(q^f_{x_j,t+T_j} - q_{x_j,t+T_j}) \tag{5.6}$$

[6] In this chapter, we perform valuation under the real-world probability measure. Therefore, the expectation in equation (5.5) is taken under the real-world probability measure and the discount rate r used in (5.4) a risk-adjusted rate (rather than a risk-free rate).

[7] If $t_b - (x_0 - T) + 1 > t + \tau - x_a$, then the expectation in $V_{L_t}(t+\tau)$ depends on the cohort effects up to $\gamma^{(4)}_{t+\tau-x_a}$ through the ARMA(R,M) process specified in equation (5.3). If $t_b - (x_0 - T) + 1 \leq t + \tau - x_a$, then the portion of the sample path beyond year-of-birth $t_b - (x_0 - T) + 1$ is irrelevant.

per \$1 notional, where x_j, T_j and $q^f_{x_j,t+T_j}$ denote the reference age, the time-to-maturity and the (fixed) forward mortality rate, respectively. Under our model assumption, $H(j,t)$ is a function of $\kappa^{(i)}_{t+T_j}$, for $i = 1, 2, 3$, and $\gamma^{(4)}_{t+T_j-x_j}$.

The forward mortality rate is determined at time t when the q-forward is launched. Following Coughlan et al. (2007), we calculate the forward mortality rate as follows:

$$q^f_{x_j,t+T_j} = (1 - T_j \times \lambda \times v_{x_j})\hat{q}_{x_j,t+T_j}, \tag{5.7}$$

where λ is the annualized Sharpe ratio demanded by the counterparty (i.e., the fixed-rate payer), $\hat{q}_{x_j,t+T_j}$ is the best estimate of $q_{x_j,t+T_j}$ that is calculated by switching off all random innovations beyond time t, and v_{x_j} is the estimated volatility of the yearly changes in the death probability at age x_j.[8]

To focus on our research objectives, we disregard population basis risk; that is, we assume that the reference population to which the q-forwards are linked is identical to the hedger's own population of individuals. However, with moderate adaptations, the proposed hedging strategy can be extended to incorporate the difference between the populations involved.

Let $H_t(j,t) = H(j,t)|\mathcal{F}_t$. It is obvious that $H_t(j,t)$ is subject to the uncertainty surrounding the following (unrealized) period effects: $\kappa^{(i)}_{t+T_j}$, where $i = 1, 2, 3$. Whether $H_t(j,t)$ is subject to cohort effect uncertainty depends on the q-forward's reference age x_j and time-to-maturity T_j. If $t + T_j - x_j > t - x_a$, then $\gamma^{(4)}_{t+T_j-x_j}$ is not yet realized and hence $H_t(j,t)$ is subject to the uncertainty surrounding $\gamma^{(4)}_{t+T_j-x_j}$. Otherwise, $\gamma^{(4)}_{t+T_j-x_j}$ is already realized and thus $H_t(j,t)$ is free from any cohort effect uncertainty.

Suppose that the longevity hedge is established at time t, where $t \geq t_b$, and that all q-forwards used are freshly launched at time t. As the longevity hedge is built to mitigate the uncertainty associated with the value of the annuity liability at time $t + \tau$, of our particular interest is

$$V_{H_t(j,t)}(t+\tau) = \mathrm{E}\left(H_t(j,t)|\mathcal{F}_{t+\tau}\right),$$

the value of the jth q-forward τ years from the time when the hedge is established. Similar to $V_{L_t}(t+\tau)$, $V_{H_t(j,t)}(t+\tau)$ is subject to the uncertainty surrounding the sample path of the period effects from years $t+1$ to $t+\min(\tau, T_j)$. If $t+T_j-x_j > t-x_a$, then $V_{H_t(j,t)}(t+\tau)$ is also subject to the uncertainty surrounding the sample path of the cohort effect over years-

[8]The forward mortality rate may also be computed using other methods, including those proposed by Chuang and Brockett (2014) and Li et al. (2011).

of-birth $t+1-x_a$ to $\min(t+\tau-x_a, t+T_j-x_j)$. We remark here that if $t+T_j-x_j \leq t-x_a$, then the jth q-forward would not be able to hedge the cohort effect uncertainty involved in the annuity liability, due to reasons that are mentioned in the previous paragraph.

5.3.3 Hedging Objectives

We consider static and dynamic hedging. For both, two hedging objectives, (1) minimization of variance and (2) minimization of Value-at-Risk, are considered.

5.3.3.1 Static Hedging

When we statically hedge the annuity liability, we build a q-forward portfolio at $t = t_b$ to mitigate the uncertainty associated with the value of the liability over a hedging horizon of $\tau > 0$ years. No adjustment is made to the portfolio during the whole τ-year period.

We use $N_j(t)$ to denote the notional amount of the jth q-forward purchased at time t. When minimizing the variance of the τ-year ahead values of the hedged position, the objective function can be expressed as

$$\min_{N_1(t_b),\ldots,N_m(t_b)} \text{Var}\left(V_{L_{t_b}}(t_b+\tau) - \sum_{j=1}^{m} N_j(t_b) V_{H_{t_b}(j,t_b)}(t_b+\tau)\right). \tag{5.8}$$

When minimizing the Value-at-Risk of the τ-year ahead values of the hedged position (relative to the unhedged mean) at the 99.5% confidence level, the objective function is given by

$$\min_{N_1(t_b),\ldots,N_m(t_b)} \text{VaR}_{0.995}\left(V_{L_{t_b}}(t_b+\tau) - V_{L_{t_b}}(t_b) - \sum_{j=1}^{m} N_j(t_b) V_{H_{t_b}(j,t_b)}(t_b+\tau)\right), \tag{5.9}$$

where $\text{VaR}_{0.995}(y)$ represents the 99.5th percentile of y.

5.3.3.2 Dynamic Hedging

To dynamically hedge the annuity liability, we first set up at time t_b a q-forward portfolio that minimizes the variance or Value-at-Risk of the values of the hedged position in $\tau = 1$ year. At time t_b+1, the existing q-forwards are closed out and a new hedge portfolio, which is composed of q-forwards that are freshly launched at time t_b+1, is built to minimize the

variance or Value-at-Risk of the hedged position in the next $\tau = 1$ year. The process is repeated over the desired duration, say Y years.

To keep the set-up simple, we assume that the q-forwards used always have reference ages $x_1, \dots x_m$ and times-to-maturity (measured from the launch date) $T_1, \dots T_m$. We further assume that no transaction cost is incurred when trading the q-forward contracts.

When we aim to minimize the variance of the 1-year ahead values of the hedged position, the objective function can be expressed as

$$\min_{N_1(t),\dots,N_m(t)} \text{Var}\left(V_{L_t}(t+1) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+1)\right), \tag{5.10}$$

for $t = t_b, t_b + 1, \dots, t_b + Y - 1$. When we aim to minimize the 99.5% Value-at-Risk of the 1-year ahead values of the hedged position (relative to the unhedged mean), the objective function is given by

$$\min_{N_1(t),\dots,N_m(t)} \text{VaR}_{0.995}\left(V_{L_t}(t+1) - V_{L_t}(t) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+1)\right), \tag{5.11}$$

for $t = t_b, t_b + 1, \dots, t_b + Y - 1$.

Note that objective function (5.11) is in line with Solvency II's one-year risk, which, as Plat (2011) summarized, consists of the following two components:

1. the risk that next year's realized mortality will be below (or above) its expectation;
2. the risk of a decrease (or increase) in expected mortality beyond next year.

Ignoring the correlations between different risk modules, the difference between $\text{VaR}_{0.995}(V_{L_t}(t+1) - V_{L_t}(t))$ (the Value-at-Risk of the unhedged position) and the Value-at-Risk minimized under (5.11) represents the reduction in the Solvency Capital Requirement (SCR) for year $t+1$ due to the q-forward hedge portfolio that is built at the beginning of the year.

5.4 Deriving the Optimal Hedging Strategies

5.4.1 Reformulating L_t and $H_t(j,t)$

As discussed in Section 5.3, L_t and $H_t(j,t)$ are functions of various unrealized period and/or cohort effects. However, these functional forms are not easy to work with, because there

exist serial correlations in the period and cohort effects. To get around this problem, we now reformulate L_t and $H_t(j,t)$ as functions of the innovations (in the stochastic processes for the period and cohort effects), which, by definition, are free of any serial correlation.

First, let us consider the period effects. Under the trivariate random walk with drift (equation (5.2)), we can express the period effects for year $t+s$ as

$$\kappa_{t+s}^{(i)} = \kappa_t^{(i)} + \mu^{(i)} s + \sum_{u=1}^{s} \eta_{t+u}^{(i)},$$

for $i = 1, 2, 3$. Given $\mathcal{F}_t$, the only random components in the expression are the following innovations: $\eta_{t+1}^{(i)}, \ldots, \eta_{t+s}^{(i)}$.

Next, we turn to the cohort effects. By rewriting the ARMA(R,M) process (equation (5.3)) into an infinite MA form, we can express the cohort effect for year-of-birth $t - x_a + s$ as

$$\gamma_{t-x_a+s}^{(4)} = \mu^{(4)} \left(1 - \sum_{i=1}^{R} \phi_i^{(4)} B^i\right)^{-1} + \left(1 + \sum_{i=1}^{M} \theta_i^{(4)} B^i\right) \left(1 - \sum_{i=1}^{R} \phi_i^{(4)} B^i\right)^{-1} \eta_{t-x_a+s}^{(4)}, \quad (5.12)$$

where B denotes the backshift operator. Given $\mathcal{F}_t$, the only random terms in the expression are the following innovations: $\eta_{t-x_a+1}^{(4)}, \ldots, \eta_{t-x_a+s}^{(4)}$.

Combining the above and the results from Section 5.3, we can draw the following conclusions. When $t_b \le t < t_b + T + \omega - x_0$ (the range of t that is of our interest), L_t is a function of the following random innovations: $\eta_{t+1}^{(i)}, \ldots \eta_{t+T+\omega-x_0}^{(i)}$, where $i = 1, 2, 3$. Also, if $t_b - (x_0 - T) + 1 > t - x_a$, then L_t is also a function of $\eta_{t-x_a+1}^{(4)}, \ldots, \eta_{t_b+T-x_0+1}^{(4)}$. All other components in L_t are non-random as of time t.

For any $t \ge t_b$, $H_t(j,t)$ is a function of the following random innovations from the period effect process: $\eta_{t+1}^{(i)}, \ldots, \eta_{t+T_j}^{(i)}$, where $i = 1, 2, 3$. If $t + T_j - x_j > t - x_a$, then $H_t(j,t)$ is also a function of $\eta_{t-x_a+1}^{(4)}, \ldots, \eta_{t+T_j-x_j}^{(4)}$. Otherwise, $H_t(j,t)$ involves no unrealized innovation from the cohort effect process. Other than the mentioned innovations, all components in $H_t(j,t)$ are non-random as of time t.

We conclude this sub-section by specifying the innovation vector $\vec{\eta}_t$ and its probability distribution. We define $\vec{\eta}_t$ as follows:

$$\vec{\eta}_t := (\eta_t^{(1)}, \eta_t^{(2)}, \eta_t^{(3)}, \eta_{t-x_a}^{(4)})'.$$

Given the model assumptions, $\vec{\eta}_t$ possesses no serial correlation. It follows a multivariate

normal distribution with a zero mean vector and a covariance matrix

$$Q^* = \begin{pmatrix} Q & 0 \\ 0 & (\sigma^{(4)})^2 \end{pmatrix},$$

where Q and $(\sigma^{(4)})^2$ are defined in Section 5.2.

5.4.2 Linear Approximations

It follows from equations (5.4) and (5.6) that both L_t and $H_t(j,t)$ are non-linear functions of the period and cohort effects (and thus non-linear functions of the innovations from the period and cohort effect processes). To enable an analytical derivation of the hedging strategies, we approximate L_t and $H_t(j,t)$ with first-order Taylor expansions around the relevant random innovations.

We let l_t and $h_t(j,t)$ be the first-order Taylor approximations of L_t and $H_t(j,t)$, respectively. We have

$$l_t = \hat{L}_t + \sum_{s=1}^{T+\omega-x_0-t} \frac{\partial L_t}{\partial \eta_{t+s}^{(i)}} \eta_{t+s}^{(i)} + \sum_{s=1}^{t_b-(x_0-T)+x_a-t+1} \frac{\partial L_t}{\partial \eta_{t-x_a+s}^{(4)}} \eta_{t-x_a+s}^{(4)}, \tag{5.13}$$

for $t_b \leq t < t_b + T + \omega - x_0$, and

$$h_t(j,t) = \hat{H}_t(j,t) + \sum_{i=1}^{3} \sum_{s=1}^{T_j} \frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(i)}} \eta_{t+s}^{(i)} + \sum_{s=1}^{T_j-x_j} \frac{\partial H_t(j,t)}{\partial \eta_{t-x_a+s}^{(4)}} \eta_{t-x_a+s}^{(4)}, \tag{5.14}$$

for $t \geq t_b$, where $\hat{L}_t$ and $\hat{H}_t(j,t)$ respectively represent the estimates of L_t and $H_{j,t}$ that are computed by setting $\vec{\eta}_{t+s}$ to $\vec{0}$ (the expected value of $\vec{\eta}_{t+s}$) for all $s > 0$. Also, throughout this chapter, all partial derivatives with respect to the innovations are evaluated at either the realized innovations (if the innovations are realized) or the expected values of the innovations (if the innovations are not yet realized).

Using equation (5.13), we obtain the following approximation for $V_{L_t}(t+\tau)$:

$$\begin{aligned} V_{L_t}(t+\tau) &\approx V_{l_t}(t+\tau) \\ &:= \mathrm{E}\left(l_t|\mathcal{F}_{t+\tau}\right) \\ &= \mathrm{E}\left(\hat{L}_t + \sum_{s=1}^{T+\omega-x_0-t} \frac{\partial L_t}{\partial \eta_{t+s}^{(i)}}\eta_{t+s}^{(i)} + \sum_{s=1}^{t_b-(x_0-T)+x_a-t+1} \frac{\partial L_t}{\partial \eta_{t-x_a+s}^{(4)}}\eta_{t-x_a+s}^{(4)} \middle| \mathcal{F}_{t+\tau}\right) \\ &= \hat{L}_t + \sum_{i=1}^{3}\sum_{s=1}^{\tau} \frac{\partial L_t}{\partial \eta_{t+s}^{(i)}}\eta_{t+s}^{(i)} + \sum_{s=1}^{\tau} \frac{\partial L_t}{\partial \eta_{t-x_a+s}^{(4)}}\eta_{t-x_a+s}^{(4)} \\ &= \hat{L}_t + \sum_{s=1}^{\tau}\left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}}\right)'\vec{\eta}_{t+s}, \end{aligned} \tag{5.15}$$

where $\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} = \left(\frac{\partial L_t}{\partial \eta_{t+s}^{(1)}}, \frac{\partial L_t}{\partial \eta_{t+s}^{(2)}}, \frac{\partial L_t}{\partial \eta_{t+s}^{(3)}}, \frac{\partial L_t}{\partial \eta_{t-x_a+s}^{(4)}}\right)'$.

Also, using equation (5.14), we have the following approximation for $V_{H_t(j,t)}(t+\tau)$:

$$\begin{aligned} V_{H_t(j,t)}(t+\tau) &\approx V_{h_t(j,t)}(t+\tau) \\ &:= \mathrm{E}\left(h_t(j,t)|\mathcal{F}_t\right) \\ &= \mathrm{E}\left(\hat{H}_t(j,t) + \sum_{i=1}^{3}\sum_{s=1}^{T_j} \frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(i)}}\eta_{t+s}^{(i)} + \sum_{s=1}^{T_j-x_j} \frac{\partial H_t(j,t)}{\partial \eta_{t-x_a+s}^{(4)}}\eta_{t-x_a+s}^{(4)} \middle| \mathcal{F}_{t+\tau}\right) \\ &= \hat{H}_t(j,t) + \sum_{i=1}^{3}\sum_{s=1}^{\tau} \frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(i)}}\eta_{t+s}^{(i)} + \sum_{s=1}^{\tau} \frac{\partial H_t(j,t)}{\partial \eta_{t-x_a+s}^{(4)}}\eta_{t-x_a+s}^{(4)} \\ &= \hat{H}_t(j,t) + \sum_{s=1}^{\tau}\left(\frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)'\vec{\eta}_{t+s}, \end{aligned} \tag{5.16}$$

where $\frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}} = \left(\frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(1)}}, \frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(2)}}, \frac{\partial H_t(j,t)}{\partial \eta_{t+s}^{(3)}}, \frac{\partial H_t(j,t)}{\partial \eta_{t-x_a+s}^{(4)}}\right)'$.

The expression for $V_{l_t}(t+\tau)$ contains the partial derivatives of L_t with respect to different innovations. Those with respect to the period-effect-related innovations can be

computed analytically as follows:

$$\begin{aligned}
\frac{\partial L_t}{\partial \eta^{(1)}_{t+k}} &= -\sum_{u=1}^{\omega-x_0} e^{-r(T+u)} \left(\prod_{s=1}^{T+u}(1-q_{x_0-T+s-1,t_b+s})\right)\left(\sum_{s=t-t_b+k}^{T+u} q_{x_0-T+s-1,t_b+s}\right); \\
\frac{\partial L_t}{\partial \eta^{(2)}_{t+k}} &= -\sum_{u=1}^{\omega-x_0} e^{-r(T+u)} \left(\prod_{s=1}^{T+u}(1-q_{x_0-T+s-1,t_b+s})\right) \\
&\quad \times \left(\sum_{s=t-t_b+k}^{T+u}(x_0-T+s-1-\bar{x})q_{x_0-T+s-1,t_b+s}\right); \\
\frac{\partial L_t}{\partial \eta^{(3)}_{t+k}} &= -\sum_{u=1}^{\omega-x_0} e^{-r(T+u)} \left(\prod_{s=1}^{T+u}(1-q_{x_0-T+s-1,t_b+s})\right) \\
&\quad \times \left(\sum_{s=t-t_b+k}^{T+u}((x_0-T+s-1-\bar{x})^2-\hat{\sigma}_x^2)q_{x_0-T+s-1,t_b+s}\right),
\end{aligned} \tag{5.17}$$

for $k = 1, \ldots, \min(\tau, w - x_0 + T + (t_b - t))$, and

$$\frac{\partial L_t}{\partial \eta^{(1)}_{t+k}} = \frac{\partial L_t}{\partial \eta^{(2)}_{t+k}} = \frac{\partial L_t}{\partial \eta^{(3)}_{t+k}} = 0,$$

for $k > w - x_0 + T + (t_b - t)$. Those with respect to the cohort-effect-related innovations can be computed analytically as follows:

$$\frac{\partial L_t}{\partial \eta^{(4)}_{t-x_a+k}} = -X(C_{t,k}) \sum_{u=1}^{\omega-x_0} e^{-r(T+u)} \left(\prod_{s=1}^{T+u}(1-q_{x_0-T+s-1,t_b+s})\right)\left(\sum_{s=t-t_b+1}^{T+u} q_{x_0-T+s-1,t_b+s}\right), \tag{5.18}$$

for $k = 1, \ldots, \tau$, where $C_{t,k} = t_b - (x_0 - T) + 1 - (t - x_a + k)$ and

$$X(s) = \begin{cases} 0, & \text{if } c < 0 \\ 1, & \text{if } s = 0 \\ \phi^{(4)}_1 + \theta^{(4)}_1, & \text{if } s = 1 \\ \sum_{i=1}^{s-1} \phi^{(4)}_i X(s-i) + (\phi^{(4)}_s + \theta^{(4)}_s), & \text{if } s \geq 2 \end{cases} \tag{5.19}$$

with $\phi^{(4)}_s = 0$ if $s > R$ and $\theta^{(4)}_s = 0$ if $s > M$.

The expression for $V_{h_t(j,t)}(t+\tau)$ contains the partial derivatives of $h_t(j,t)$ with respect to different innovations. Those respect to the period-effect-related innovations can be

calculated analytically as follows:

$$\begin{array}{rcl} \frac{\partial H_j(j,t)}{\partial \eta^{(1)}_{t+k}} & = & -e^{-rT_j}(1-q_{x_j,t+T_j})q_{x_j,t+T_j}; \\ \frac{\partial H_j(j,t)}{\partial \eta^{(2)}_{t+k}} & = & -e^{-rT_j}(x_j-\bar{x})(1-q_{x_j,t+T_j})q_{x_j,t+T_j}; \\ \frac{\partial H_j(j,t)}{\partial \eta^{(3)}_{t+k}} & = & -e^{-rT_j}((x_j-\bar{x})^2-\hat{\sigma}^2_x)(1-q_{x_j,t+T_j})q_{x_j,t+T_j}, \end{array} \tag{5.20}$$

for $k = 1, \ldots, \min(\tau, T_j)$, and

$$\frac{\partial H_j(j,t)}{\partial \eta^{(1)}_{t+k}} = \frac{\partial H_j(j,t)}{\partial \eta^{(2)}_{t+k}} = \frac{\partial H_j(j,t)}{\partial \eta^{(3)}_{t+k}} = 0,$$

for $k > T_j$. Those with respect to the cohort-effect-related innovations can be calculated analytically as follows:

$$\frac{\partial H_j(j,t)}{\partial \eta^{(4)}_{t-x_a+k}} = -X(T_j - x_j + x_a - k)e^{-rT_j}(1-q_{x_j,t+T_j})q_{x_j,t+T_j}. \tag{5.21}$$

5.4.3 Minimizing Variance

As discussed in Section 5.3.3, the variance to be minimized is

$$\text{Var}\left(V_{L_t}(t+\tau) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+\tau)\right),$$

where $t = t_b$ for a static hedge, and $t = t_b, \ldots, T$ and $\tau = 1$ for a dynamic hedge.

Using equations (5.15) and (5.16), the variance to be minimized can be approximated as

$$\begin{array}{rl} & \text{Var}\left(V_{l_t}(t+\tau) - \sum_{j=1}^{m} N_j(t) V_{h_t(j,t)}(t+\tau)\right) \\ = & \sum_{s=1}^{\tau} \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right). \end{array} \tag{5.22}$$

We can calculate the optimal notional amounts $\hat{N}_1(t), \ldots, \hat{N}_m(t)$ by solving the first-order conditions, that is, by setting the first partial derivatives of (5.22) with respect

to $N_1(t), \ldots, N_m(t)$ to zero. The solution is given by

$$\begin{pmatrix} \hat{N}_1(t) \\ \vdots \\ \hat{N}_m(t) \end{pmatrix} = \begin{pmatrix} \sum_{s=1}^{\tau} (\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}} & \cdots & \sum_{s=1}^{\tau} (\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial H_t(m,t)}{\partial \vec{\eta}_{t+s}} \\ \vdots & \ddots & \vdots \\ \sum_{s=1}^{\tau} (\frac{\partial H_t(m,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}} & \cdots & \sum_{s=1}^{\tau} (\frac{\partial H_t(m,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial H_t(m,t)}{\partial \vec{\eta}_{t+s}} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{s=1}^{\tau} (\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial L_t}{\partial \vec{\eta}_{t+s}} \\ \vdots \\ \sum_{s=1}^{\tau} (\frac{\partial H_t(m,t)}{\partial \vec{\eta}_{t+s}})' Q^* \frac{\partial L_t}{\partial \vec{\eta}_{t+s}} \end{pmatrix}. \tag{5.23}$$

The second order conditions (not shown) indicate that the solution above minimizes the variance specified in equation (5.22). Note that the solution above does not depend on the forward mortality rate, and hence it is unaffected by the assumption about the market price of risk (λ).

To develop a deeper insight into the solution, let us focus on the case when $m = 1$ q-forward is used. In this case, the optimal notional amount can be expressed as

$$\hat{N}_1(t) = \frac{k_2}{k_1}, \tag{5.24}$$

and the minimized variance is

$$k_3 - \frac{k_2^2}{k_1}, \tag{5.25}$$

where

$$k_1 = \sum_{s=1}^{\tau} \left(\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}} \right)' Q^* \left(\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}} \right) \tag{5.26}$$

is the approximate variance of $H_t(1,t)$,

$$k_2 = \sum_{s=1}^{\tau} \left(\frac{\partial H_t(1,t)}{\partial \vec{\eta}_{t+s}} \right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} \right) \tag{5.27}$$

is the approximate covariance between $H_t(1,t)$ and L_t, and

$$k_3 = \sum_{s=1}^{\tau} \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} \right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} \right) \tag{5.28}$$

is the approximate variance of L_t.

How would the solution change if relevant cohort effect uncertainty is not taken into account? Suppose that cohort effect does exist but the hedger mistakenly ignores cohort effect uncertainty (i.e., treating $(\sigma^{(4)})^2$ as zero). In this hypothetical situation, the expression of the optimized notional amount would exclude the terms involving $(\sigma^{(4)})^2$ and become

$$\hat{\mathcal{N}}_1(t) = \frac{k_2 - c_2}{k_1 - c_1} = \frac{k_2\left(1 - \frac{c_2}{k_2}\right)}{k_1\left(1 - \frac{c_1}{k_1}\right)},$$

where

$$c_1 = \sum_{s=1}^{\tau} \left(\frac{\partial H_t(j,t)}{\partial \eta^{(4)}_{t-x_a+s}}\right)^2 (\sigma^{(4)})^2$$

can be viewed as the portion of the variance of $H_j(j,t)$ that is contributed from the innovations associated with the cohort effects, and

$$c_2 = \sum_{s=1}^{\tau} \frac{\partial H_t(j,t)}{\partial \eta^{(4)}_{t-x_a+s}} \cdot \frac{\partial L_t}{\partial \eta^{(4)}_{t-x_a+s}} (\sigma^{(4)})^2$$

can be understood as the portion of the covariance between $H_t(j,t)$ and L_t that is contributed from the innovations associated with the cohort effects. It is clear that $\hat{\mathcal{N}}_1(t)$ is either larger or smaller than the actual optimal notional amount $\hat{N}_1(t)$ (which incorporates cohort effect uncertainty) unless $c_1/k_1 = c_2/k_2$. It can also be seen that $\hat{\mathcal{N}}_1(t)$ is close to $\hat{N}_1(t)$ when c_1/k_1 and c_2/k_2 are close to zero. The conditions under which under which $\hat{\mathcal{N}}_1(t)$ and $\hat{N}_1(t)$ are equal (or close to each other) are further discussed in Section 5.6 when we present the numerical results. Substituting $N_1(t) = \hat{\mathcal{N}}_1(t)$ into equation (5.22), we obtain the variance of the values of the hedged position when cohort effect uncertainty is ignored in the calibration procedure but it actually exists:

$$k_3 - 2k_2\frac{k_2 - c_2}{k_1 - c_1} + k_1\left(\frac{k_2 - c_2}{k_1 - c_1}\right)^2. \tag{5.29}$$

The difference between expressions (5.29) and (5.25) is

$$\frac{1}{k_1}\left(k_2 - k_1\frac{k_2 - c_2}{k_1 - c_1}\right)^2,$$

which is straightly greater than zero unless $c_1/k_1 = c_2/k_2$.[9] It follows that ignoring cohort effect uncertainty generally leads to a sub-optimal variance minimization.

5.4.4 Minimizing Value-at-Risk

As discussed in Section 5.3.3, the Value-at-Risk to be minimized is

$$\mathrm{VaR}_{0.995}\left(V_{L_t}(t+\tau) - V_{L_t}(t) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+\tau)\right),$$

which can be approximated as

$$\mathrm{VaR}_{0.995}\left(V_{l_t}(t+\tau) - V_{L_t}(t) - \sum_{j=1}^{m} N_j(t) V_{h_t(j,t)}(t+\tau)\right),$$

where $t = t_b$ for a static hedge, and $t = t_b, \ldots, T$ and $\tau = 1$ for a dynamic hedge.

On the basis of equations (5.15) and (5.16) and the fact that the innovation vectors are i.i.d. multivariate-normally distributed, we know that

$$V_{l_t}(t+\tau) - V_{L_t}(t) - \sum_{j=1}^{m} N_j(t) V_{h_t(j,t)}(t+\tau)$$

follows a normal distribution with a mean of

$$\mu_t = -\sum_{j=1}^{m} N_j(t) \hat{H}_t(j,t) \tag{5.30}$$

and a variance of

$$\Sigma_t = \sum_{s=1}^{\tau} \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right). \tag{5.31}$$

As a shorthand, we use $\pi_{0.995}$ to denote the Value-at-Risk at the 99.5% significance

[9] We can interpret k_1 as the approximate variance of $H_j(j,t)$, and thus it is straightly greater than zero.

level. By definition,

$$\Pr\left(V_{l_t}(t+\tau) - V_{l_t}(t) - \sum_{j=1}^{m} N_j(t) V_{h_t(j,t)}(t+\tau) \leq \pi_{0.995}\right) = 0.995.$$

Rearranging, we have

$$\Pr\left(\frac{V_{l_t}(t+\tau) - V_{l_t}(t) - \sum_{j=1}^{m} N_j(t) V_{h_t(j,t)}(t+\tau) - \mu_t}{\sqrt{\Sigma_t}} \leq \frac{\pi_{0.995} - \mu_t}{\sqrt{\Sigma_t}}\right) = 0.995,$$

which gives

$$\pi_{0.995} = \mu_t + \sqrt{\Sigma_t} \cdot \Phi^{-1}(0.995), \tag{5.32}$$

where $\Phi^{-1}(0.995)$ is the 99.5% quantile of the standard normal distribution.

To derive the optimal notional amounts, we take partial derivative of $\pi_{0.995}$ with respect to $N_1(t), \ldots, N_m(t)$:

$$\frac{\partial \pi_{0.995}}{\partial N_i(t)} = -\hat{H}_t(i,t) - \frac{\Phi^{-1}(0.995) \cdot \sum_{s=1}^{\tau} \left(\frac{\partial H_t(i,t)}{\partial \vec{\eta}_{t+s}}\right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)}{\sqrt{\sum_{s=1}^{\tau} \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)' Q^* \left(\frac{\partial L_t}{\partial \vec{\eta}_{t+s}} - \sum_{j=1}^{m} N_j(t) \frac{\partial H_t(j,t)}{\partial \vec{\eta}_{t+s}}\right)}},$$

for $i = 1, \ldots, m$. The optimal notional amounts $\hat{N}_1(t), \ldots, \hat{N}_m(t)$ are calculated by solving the first order conditions,

$$\frac{\partial \pi_{0.995}}{\partial N_i(t)}_{|N_i(t)=\hat{N}_i(t)} = 0, \quad i = 1, \ldots, m$$

and verifying the second order conditions.

When $m = 1$ q-forward is used, the first-order condition boils down to a quadratic equation. The root that satisfies the second-order condition

$$\frac{\partial^2 \pi_{0.995}}{\partial (N_i(t))^2}_{|N_1(t)=\hat{N}_1(t)} > 0$$

is

$$\hat{N}_1(t) = \frac{-b - \sqrt{b^2 - 4ac}}{2a},$$

where

$$a = \left(\Phi^{-1}(0.995) \cdot k_1\right)^2 - (\hat{H}_t(1,t))^2 \cdot k_1,$$

$$b = 2\left(-(\Phi^{-1}(0.995))^2 k_1 + (\hat{H}_t(1,t))^2\right) \cdot k_2,$$

$$c = (\Phi^{-1}(0.995))^2 (k_2)^2 - (\hat{H}_t(1,t))^2 k_3,$$

and k_1, k_2 and k_3 are defined in Section 5.4.3.

It is noteworthy that $\hat{N}_1(t)$ exists if and only if $b^2 \geq 4ac$. It can be shown that

$$b^2 - 4ac = 4k_1k_3\left(\frac{k_2^2}{k_1k_3} - 1\right)(\hat{H}_t(1,t))^2((\hat{H}_t(1,t))^2 - (\Phi^{-1}(0.995))^2 k_1).$$

Both k_1 and k_3 are positive, as they respectively represent the approximate variances of $H_1(1,t)$ and L_t. Also, it follows from the interpretations of k_1, k_2 and k_3 that $k_2/(k_1k_3)$ represents the square of the (approximate) correlation coefficient between $V_{l_t}(t+\tau)$ and $V_{h_t(1,t)}(t+\tau)$, and is thus strictly smaller than 1. Consequently, a sufficient condition for $b^2 \geq 4ac$ is

$$(\hat{H}_t(1,t))^2 \leq (\Phi^{-1}(0.995))^2 k_1, \tag{5.33}$$

which means $(\hat{H}_t(1,t))^2$ has to be sufficiently small. On the basis of the pricing formula specified in equation (5.7), we have

$$\hat{H}_t(j,t) = e^{-rT_j}((1 - T_j \times \lambda \times v_{x_j})\hat{q}_{x_j,t+T_j} - \hat{q}_{x_j,t+T_j}) = -(T_j \times \lambda \times v_{x_j})\hat{q}_{x_j,t+T_j}. \tag{5.34}$$

Condition (5.33) therefore means that the Sharpe ratio demanded by the q-forward's counterparty cannot be too large. This condition is studied more deeply in Section 5.6.3 where we examine the sensitivity of the hedging results to the assumed value of λ.

It is also interesting to note that if the q-forwards are costless (i.e., the Sharpe ratio λ demanded by the counterparty is zero), then $\hat{H}_t(j,t) = 0$ for $j = 1, \ldots, m$.[10] In this case, the optimal notional amounts obtained by minimizing Value-at-Risk and minimizing

[10]When $\lambda = 0$, the forward mortality rate $q^f_{x_j,t+T_j}$ equals $\hat{q}_{x_j,t+T_j}$, the best estimate of $q_{x_j,t+T_j}$ calculated by switching off all random innovations beyond time t. As such, $H_t(j,t) = e^{-rT_j}(\hat{q}_{x_j,t+T_j} - q_{x_j,t+T_j})$ and

variance would be identical. In particular, for $m = 1$, when $\hat{H}_t(j,t) = 0$, we have

$$a = (\Phi^{-1}(0.995)k_1)^2, \quad b = -2(\Phi^{-1}(0.995))^2 k_1 k_2, \quad c = (\Phi^{-1}(0.995)k_2)^2$$

and consequently $\hat{N}_1(t) = k_2/k_1$, which is the same as the expression for $\hat{N}_1(t)$ when the objective is to minimize variance.

5.5 Evaluating Hedge Effectiveness

5.5.1 The Metrics

When the hedging objective is to minimize variance, we measure hedge effectiveness as the proportion of variance reduced. For a static hedge, the metric can be expressed as

$$\text{HE}^{\text{Var}} = 1 - \frac{\text{Var}(V_{L_{t_b}}(t_b+\tau) - \sum_{j=1}^{m} N_j(t_b) V_{H_{t_b}(j,t_b)}(t_b+\tau))}{\text{Var}(V_{L_{t_b}}(t_b+\tau))}. \tag{5.35}$$

For a dynamic hedge, the metric (applicable to year $t+1$) is given by

$$\text{HE}_t^{\text{Var}} = 1 - \frac{\text{Var}(V_{L_t}(t+1) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+1))}{\text{Var}(V_{L_t}(t+1))}, \tag{5.36}$$

for $t = t_b, t_b+1, \ldots, t_b+Y-1$. The value of this metric is close to 1 if the longevity hedge is effective, and close to 0 if otherwise.

When the hedging objective to minimize Value-at-Risk, we measure hedge effectiveness as the absolute reduction in Value-at-Risk. For a static hedge, the metric is given by

$$\begin{aligned} \text{HE}^{\text{VaR}_{0.995}} &= \text{VaR}_{0.995}\left(V_{L_{t_b}}(t+\tau) - V_{L_{t_b}}(t_b)\right) \\ &\quad - \text{VaR}_{0.995}\left(V_{L_{t_b}}(t+\tau) - V_{L_{t_b}}(t_b) - \sum_{j=1}^{m} N_j(t_b) V_{H_{t_b}(j,t_b)}(t_b+\tau)\right). \end{aligned} \tag{5.37}$$

For a dynamic hedge, the metric (applicable to year $t+1$) can be expressed as:

$$\begin{aligned} \text{HE}_t^{\text{VaR}_{0.995}} &= \text{VaR}_{0.995}\left(V_{L_t}(t+1) - V_{L_t}(t)\right) \\ &\quad - \text{VaR}_{0.995}\left(V_{L_t}(t+1) - V_{L_t}(t) - \sum_{j=1}^{m} N_j(t) V_{H_t(j,t)}(t+1)\right), \end{aligned} \tag{5.38}$$

$\hat{H}_t(j,t) = e^{-rT_j}(\hat{q}_{x_j,t+T_j} - \hat{q}_{x_j,t+T_j}) = 0.$

for $t = t_b, t_b + 1, \ldots, t_b + Y - 1$. Of course, the larger this metric is, the more effective the longevity hedge is. Also, as discussed in Section 5.3.3.2, the metric specified by equation (5.38) can be loosely interpreted to mean the reduction in the Solvency II SCR for year $t + 1$ due to the q-forward hedge portfolio that is built at the beginning of the year.

5.5.2 Evaluation by Analytical Approximations

The metrics specified in the previous subsection can be computed using analytical approximations. In particular, we can use equation (5.22) to approximately calculate metrics (5.35) and (5.36), and equations (5.30), (5.31) and (5.32) to approximately compute metrics (5.37) and (5.38). This method requires no simulation.

5.5.3 Evaluation by Simulations

5.5.3.1 The Potential Computational Burden

Alternatively, for a more exact calculation of the hedge effectiveness metrics, we may use simulations. However, this approach to evaluating a value hedge typically requires nested simulations.

To illustrate, let us consider a static hedge with a hedging horizon of τ. Suppose that M_1 sets of sample paths (of the period effects beyond time t_b and the cohort effects beyond year-of-birth $t_b - x_a$) are generated for assessing the effectiveness of the hedge. For each one of the M_1 sets of sample paths, we need to value the annuity liability and q-forwards as of time $t_b + \tau$; that is, to calculate $V_{L_{t_b}}(t_b + \tau) = \mathrm{E}(L_{t_b} | \mathcal{F}_{t_b+\tau})$ and $V_{H_{t_b}(j,t_b)}(t_b + \tau) = \mathrm{E}(H_{t_b}(j, t_b) | \mathcal{F}_{t_b+\tau})$, which have no exact analytical form. To evaluate these expectations, another M_2 sets of sample paths (of the period effects beyond time $t_b + \tau$ and the cohort effects beyond year-of-birth $t_b - x_a + \tau$) have to be generated. Because of the serial correlations in the period and cohort effect processes, these expectations depend on the realizations from time $t_b + 1$ to $t_b + \tau$ and from year-of-birth $t_b - x_a + 1$ to $t_b - x_a + \tau$. What this means is that we need to generate additional M_2 sets of sample paths for each one of the M_1 sets of sample paths. In effect, $M_1 \times M_2$ sets of sample paths are required. Generally speaking, M_1 and M_2 are large, say 10,000, and therefore the computation burden is substantial. If the hedge is dynamic, then the computation burden would be even more significant, requiring $M_1 \times M_2 \times Y$ sets of sample paths.

$$\begin{array}{ccc|ccc}
\vec{\eta}_{t_b+1}(1) & \cdots & \vec{\eta}_{t_b+\tau}(1) & \vec{\eta}_{t_b+\tau+1}(1) & \cdots & \vec{\eta}_{t_b+T+\omega-x_0}(1) \\
\vec{\eta}_{t_b+1}(2) & \cdots & \vec{\eta}_{t_b+\tau}(2) & \vec{\eta}_{t_b+\tau+1}(2) & \cdots & \vec{\eta}_{t_b+T+\omega-x_0}(2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vec{\eta}_{t_b+1}(M_1) & \cdots & \vec{\eta}_{t_b+\tau}(M_1) & \vec{\eta}_{t_b+\tau+1}(M_1) & \cdots & \vec{\eta}_{t_b+T+\omega-x_0}(M_1)
\end{array}$$

Figure 5.4: An illustration of the innovation-based simulation method. In the diagram, $\vec{\eta}_u(i)$ represents the innovation vector at $t = u$ in the ith sample path. The block enclosed by the rectangle is used for computing $V_{L_{t_b}}(t_b + \tau)$ and $V_{H_{t_b}(j,t_b)}(t_b + \tau)$ for each node i at $t = \tau$.

5.5.3.2 An Innovation-Based Simulation Procedure

We now propose a method that saves the need for nested simulations. As discussed in Section 5.4.1, L_t and $H_t(j, t)$ can be expressed as functions in which the innovation vectors $\vec{\eta}_{t+s}$, $s = 1, 2, \ldots$, are the only random components. On the basis of this fact, in the proposed method we simulate sample paths of the innovations vectors rather than the period and cohort effects. We can then use the fact that the innovation vectors are not serially correlated to save computational effort.

To explain, we consider again a static hedge with a hedging horizon of τ. Suppose that we generate M_1 sample paths of innovation vectors to evaluate the effectiveness of the hedge, and for each of these M_1 sample paths, we use M_2 sample paths of innovation vectors to value the liability and the hedge portfolio at time $t_b + \tau$. Because the innovation vectors possess no serial correlation, the innovation vectors used for valuation at $t = t_b + \tau$ do not depend on the realized innovation vectors at $t = t_b + 1, \ldots, t_b + \tau$. Hence, the same set of M_2 sample paths for valuation purposes can be used for each of the M_1 sample paths, thereby avoiding the need for nested simulations.

The idea is illustrated in Figure 5.4. For ease of exposition, we assume here that $M_2 = M_1$. As shown in the diagram, M_1 sample paths of the innovation vectors for $t = t_b + 1, \ldots, t_b + T + \omega - x_0$ are generated. We can use these M_1 sample paths to compute $V_{L_{t_b}}(t_b)$ in metrics (5.35) and (5.37). Because the innovation vectors are not serially correlated, $\vec{\eta}_{t_b+s}$ for $s > \tau$ do not depend on the realizations of $\vec{\eta}_{t_b+s}$ for $s \leq \tau$. As such, we can use the same collection of innovation vectors to calculate $V_{L_{t_b}}(t_b + \tau)$ (the liability's value) and $V_{H_{t_b}(j,t_b)}(t_b + \tau)$ (the hedge portfolio's value) in metrics (5.35) and (5.37) for each note $i = 1, \ldots, M_1$ at time $t_b + \tau$. Furthermore, because the innovation vectors are identically distributed, we can 'reuse' part of the M_1 simulated innovation vectors (the portion enclosed by the rectangle) to compute $V_{L_{t_b}}(t_b+\tau)$ and $V_{H_{t_b}(j,t_b)}(t_b+\tau)$. In the entire process of hedge effectiveness evaluation, only M_1 sample paths of innovation vectors are needed.

By dividing the innovation vectors shown in Figure 5.4 into (overlapping) blocks with the same height (M_1) but different widths (ranging from 1 to $T + \omega - x_0 - 1$), the idea can also be applied to an annually adjusted dynamic hedge. The total number of sample paths required is still M_1, even if the hedge is rebalanced periodically.

5.5.4 The Best Achievable Hedge Effectiveness

To assess the quality of the proposed method for calibrating longevity hedges, we compare the hedge effectiveness resulting from the proposed method against the best achievable hedge effectiveness, which is calculated as follows:

(i) treat the M_1 simulated mortality scenarios as known a priori outcomes;

(ii) on the basis of the M_1 scenarios, numerically identify the notional amounts that would optimize the chosen hedge effectiveness metric;

(iii) the hedge effectiveness resulting from the notional amounts identified in the previous step can be interpreted as the best achievable hedge effectiveness.

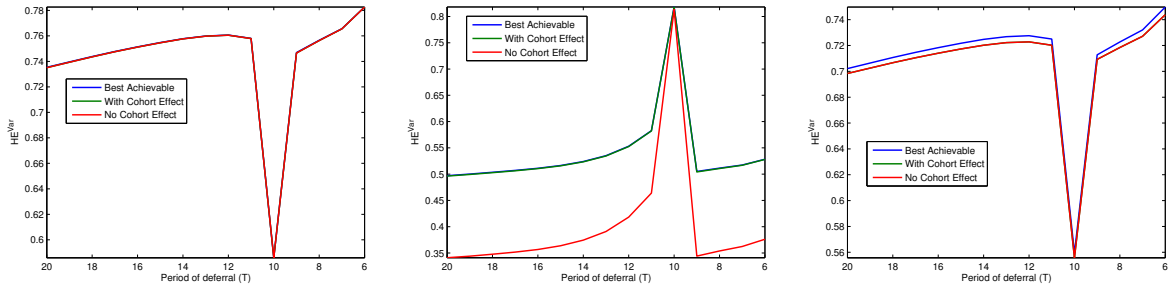
Of course, if the proposed calibration method is good, then it should yield a hedge effectiveness that is sufficiently close to the best achievable hedge effectiveness.

5.6 Numerical Illustrations

5.6.1 The Baseline Results

Let us begin by stating the baseline assumptions. In line with the set-up described in Section 5.3, the liabilities being hedged are deferred life annuities that are sold at the end of year $t_b = 2011$ when the data sample period terminates. For all annuities, the age at which the period of deferral ends is always $x_0 = 60$, and the highest attainable age is $\omega = 100$. The annuities have deferral periods ranging from $T = 6$ to $T = 20$ years, which means that they are respectively associated with years-of-birth from 1958 to 1972. As the largest year-of-birth covered by the data sample is $t_b - x_a = 2011 - 50 = 1961$, some of the annuity liabilities are subject to cohort effect uncertainty but some are not.

The hedge portfolio consists of $m = 1$ q-forward. It is assumed that three q-forwards with the same reference age ($x_1 = 65$) but different times-to-maturity ($T_1 = 11, 16, 21$)



(a) $x_1 = 65$, $T_1 = 11$ (b) $x_1 = 65$, $T_1 = 16$ (c) $x_1 = 65$, $T_1 = 21$

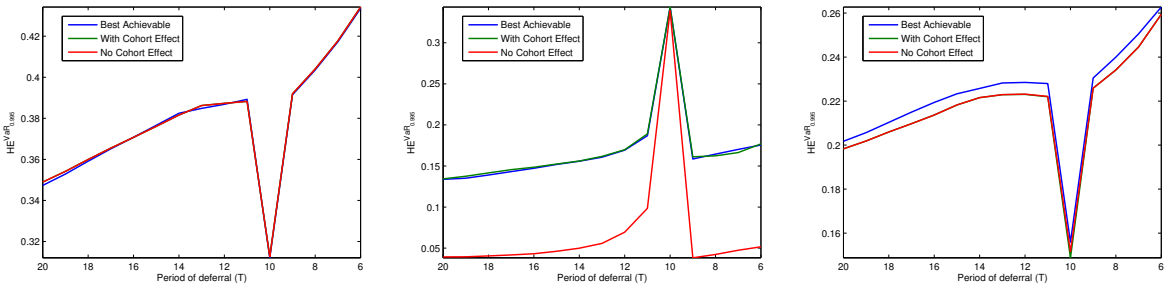
Figure 5.5: The relationship between HE^{Var} and the period of deferral T for different q-forward choices.

are available in the market. These combinations of x_1 and T_1 imply that the q-forwards are associated with years-of-birth 1957, 1962 and 1967, respectively. When calculating the forward mortality rates, it is assumed that the annualized Sharpe ratio demanded by the counterparty is $\lambda = 0$.

The hedge is established at the end of year $t_b = 2011$ (i.e., at the same time when the annuities are sold). We consider a hedging horizon of $\tau = 1$ year for the following two reasons. First, the results for $\tau = 1$ are the most informative, as a dynamic hedge is simply a repeated execution of static hedges with $\tau = 1$. Second, when $\tau = 1$, some of the analytical expressions in Section 5.4 can be simplified, thereby allowing us to more easily link the empirical results to the theoretical work. When valuing the annuity liabilities and the q-forwards, a continuously compounded risk-adjusted discount rate of $r = 0.01$ per annum is assumed.

The relationships between the calculated hedge effectiveness and the annuity's period of deferral under different circumstances are displayed in Figure 5.5 (when the objective is to minimize variance) and Figure 5.6 (when the objective is to minimize Value-at-Risk). Separate panels are used to distinguish the results based on different q-forward specifications. Each panel shows three sets of results that are obtained from the same collection of sample paths: (1) the results that are based on the proposed calibration method with cohort effect uncertainty taken into consideration (the green line labeled "with cohort effect"); (2) the results that are based on the proposed calibration method but cohort effect uncertainty is ignored (the red line labeled "no cohort effect"); (3) the best achievable results given the collection of sample paths (the blue line labeled "best achievable"). All results are calculated using the innovation-based simulation procedure described in Section 5.5.3.

Let us first compare the green and blue lines. For all circumstances under consideration,



(a) $x_1 = 65$, $T_1 = 11$ (b) $x_1 = 65$, $T_1 = 16$ (c) $x_1 = 65$, $T_1 = 21$

Figure 5.6: The relationship between $\text{HE}^{\text{VaR}_{0.995}}$ and the period of deferral T for different q-forward choices.

these two lines are very close to each other, indicating that the proposed calibration method yields a hedge effectiveness that is very close to the highest hedge effectiveness that one can possibly achieve.

We then compare the green and red lines. The gap between these two lines reflects the benefit of incorporating cohort effect uncertainty in the calibration procedure. When the q-forward with $x_1 = 65$ and $T_1 = 11$ is used (the left panels), the green and red lines coincide. This outcome is because the q-forward, which is linked to a cohort (year-of-birth 1957) that is already realized in the data sample, is not subject to any cohort effect uncertainty. As a result, no matter how the hedge is calibrated, it is not able to offset any cohort effect uncertainty involved in the annuity liability. The other two q-forwards are related to unrealized cohorts (years-of-birth 1962 and 1967), so a gap between the green and red lines exists when either one of them is used. For the q-forward with $x_1 = 65$ and $T_1 = 16$ (the middle panels), the benefit of incorporating cohort effect uncertainty is remarkable, but for the q-forward with $x_1 = 65$ and $T_1 = 21$ (the right panels), the benefit is only marginal.

We now explain why the benefit of incorporating cohort effect uncertainty depends on the choice of q-forwards. Recall that for a static hedge with $m = 1$ q-forward, the optimized notional amount[11] when cohort effect uncertainty is ignored is

$$\hat{\mathcal{N}}_1(t_b) = \frac{k_2\left(1 - \frac{c_2}{k_2}\right)}{k_1\left(1 - \frac{c_1}{k_1}\right)},$$

[11]We assume $\lambda = 0$ in the baseline results. As discussed in Section 5.4.4, when $\lambda = 0$, the optimal notional amounts for both hedging objectives under consideration would be the same.

whereas the optimized notional amount when cohort effect uncertainty is taken into account is

$$\hat{N}_1(t_b) = \frac{k_2}{k_1}.$$

These two notional amounts (and hence the corresponding hedge effectiveness values) are close to each other if

$$\frac{c_1}{k_1} = \frac{\left(\frac{\partial H_{t_b}(1,t_b)}{\partial \eta^{(4)}_{t_b-x_a+1}}\right)^2 \left(\sigma^{(4)}\right)^2}{\left(\frac{\partial H_{t_b}(1,t_b)}{\partial \vec{\eta}_{t_b+1}}\right)' Q^* \left(\frac{\partial H_{t_b}(1,t_b)}{\partial \vec{\eta}_{t_b+1}}\right)}, \tag{5.39}$$

the proportion of the variance of $H_{t_b}(1,t_b)$ that is contributed from cohort effect uncertainty, and

$$\frac{c_2}{k_2} = \frac{\left(\frac{\partial H_{t_b}(1,t_b)}{\partial \eta^{(4)}_{t_b-x_a+1}}\right)\left(\frac{\partial L_{t_b}}{\partial \eta^{(4)}_{t_b-x_a+1}}\right)\left(\sigma^{(4)}\right)^2}{\left(\frac{\partial L_{t_b}}{\partial \vec{\eta}_{t_b+1}}\right)' Q^* \left(\frac{\partial H_{t_b}(1,t_b)}{\partial \vec{\eta}_{t_b+1}}\right)}, \tag{5.40}$$

the proportion of the covariance between $H_{t_b}(1,t_b)$ and L_{t_b} that is contributed from cohort effect uncertainty, are close to zero. Equivalently speaking, the benefit of incorporating cohort effect uncertainty tends to be moderate if cohort effects are not that influential in $\text{Var}(H_{t_b}(1,t_b))$ and $\text{cov}(H_{t_b}(1,t_b), L_{t_b})$.

Table 5.2 reports the values of c_1/k_1 and c_2/k_2 for different combinations of the q-forward's maturity (T_1) and the liability's deferral period (T) when the q-forward's reference age x_1 is fixed to 65. As expected, for $T_1 \leq 15$, both ratios are zero because in this case $H_{t_b}(1,t_b)$ is not related to any unrealized cohort. Also, when $T = 5$, $c_2/k_2 = 0$ for all values of T_1 because in this case L_t is not related to any realized cohort. Other than c_2/k_2 for $T = 5$, the ratios are the highest when $T_1 = 16$, but reduce rapidly as T_1 increases. This pattern explains why the gaps between the red and green lines in the middle panels of Figures 5.5 and 5.6 are wide, but those in the right panels are narrow.

A natural question to ask is what attributes to the observed patterns of c_1/k_1 and c_2/k_2. For a given specification of the liability being hedged, c_1/k_1 and c_2/k_2 tend to be small when

$$\frac{\partial H_1(1,t_b)}{\partial \eta^{(4)}_{t_b-x_a+1}} = -X(T_j - x_j + x_a - 1)e^{-rT_1}(1 - q_{x_1,t_b+T_1})q_{x_1,t_b+T_1}$$

Table 5.2: The values of c_1/k_1 and c_2/k_2 for different combinations of the q-forward's maturity (T_1) and the liability's deferral period (T). The q-forward's reference age x_1 is fixed to 65.

T_1	≤ 15	16	17	18	19	20	21
				$T = 20$			
c_1/k_1	0	0.3336	0.0115	0.0036	0.0012	0.0004	0.0001
c_2/k_2	0	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
				$T = 15$			
c_1/k_1	0	0.3336	0.0115	0.0036	0.0012	0.0004	0.0001
c_2/k_2	0	0.0065	0.0010	0.0006	0.0003	0.0002	0.0001
				$T = 10$			
c_1/k_1	0	0.3336	0.0115	0.0036	0.0012	0.0004	0.0001
c_2/k_2	0	0.2995	0.0612	0.0353	0.0201	0.0114	0.0064
				$T = 5$			
c_1/k_1	0	0.3336	0.0115	0.0036	0.0012	0.0004	0.0001
c_2/k_2	0	0	0	0	0	0	0

is small in magnitude. It can be seen that the partial derivative is proportional to $X(T_1 - x_1 + x_a - 1)$. Specified in equation (5.19), $X(s)$ is the coefficient of $\eta^{(4)}_{t-x_a+s}$ in equation (5.12), the infinite MA form of the cohort effect process. Given the estimates of $\theta^{(4)}_1$ and $\phi^{(4)}_1$, $X(T_1 - x_1 + x_a - 1) = X(T_1 - 65 + 50 + 1)$ equals 1 when $T_1 = 16$; however, it reduces dramatically to 0.1524 when $T_1 = 17$ and reduces in an exponential manner to 0 when T_1 increases further. Thus, the assumed cohort effect process has a significant influence on c_1/k_1 and c_2/k_2, and consequently the benefit of incorporating cohort effect uncertainty.

The results shown in Figures 5.5 and 5.6 highlight an important fact: even if the annuity liability is not linked to any unrealized cohort, incorporating cohort effect uncertainty into the calibration process may still result in a more effective hedge. This fact can be identified most easily in the middle panels, from which we observe that the gaps between the green and red lines remain when the annuity's period of deferral T is less than 10 (i.e., the year-of-birth associated with the annuity is earlier than 1961). We can explain this fact by considering again the expressions for $\hat{\mathcal{N}}_1(t_b)$ and $\hat{N}_1(t_b)$. When the annuity liability is not linked to any unrealized cohort, c_2 equals zero. However, c_1 may still be positive (and large in magnitude) as it is not related to the annuity liability. Therefore, $\hat{\mathcal{N}}_1(t_b)$ may still be significantly different from $\hat{N}_1(t_b)$, which means the incorporation of cohort effect uncertainty still matters.

In each panel of Figures 5.5 and 5.6, we observe a spike at $T = 10$. More interestingly,

the spikes in the middle panels point upwards, but the spikes in the left and right panels point downwards. We can explain the spikes in the patterns of HE^{Var} by noting that the analytical approximation of the optimized HE^{Var} is given by

$$1 - \frac{k_3 - \frac{k_2^2}{k_1}}{k_3} = \frac{k_2^2}{k_1 k_3}, \tag{5.41}$$

in which k_1 is positively related to $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1}$, k_3 is positively related to $\partial L_{t_b}/\partial \eta^{(4)}_{t_b-x_a+1}$, and k_2 is positively related to the *product* of $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1}$ and $\partial L_{t_b}/\partial \eta^{(4)}_{t_b-x_a+1}$.[12] Furthermore, $\partial H_{t_b}(1,t_b)/\eta^{(4)}_{t_b-x_a+1}$ is proportional to $X(T_1 - x_1 + x_a - 1)$, and

$$\frac{\partial L_t}{\partial \eta^{(4)}_{t_b-x_a+1}} = -X(x_a - x_0 + T) \sum_{u=1}^{\omega-x_0} e^{-r(T+u)} \left(\prod_{s=1}^{T+u} (1 - q_{x_0-T+s-1,t_b+s}) \right) \left(\sum_{s=t-t_b+1}^{T+u} q_{x_0-T+s-1,t_b+s} \right)$$

is proportional to $X(x_a - x_0 + T)$. The following conclusions can then be drawn.

- $x_1 = 65$ and $T_1 = 11$ (the left panel)

  At $T = 10$, we have $X(x_a - x_0 + T) = X(50 - 60 + 10) = X(0)$. As mentioned earlier, given our estimates of $\theta_1^{(4)}$ and $\phi_1^{(4)}$, $X(s)$ is exceptionally large when $s = 0$, and therefore when T approaches 10, k_3 (which is positively related to $X(x_a - x_0 + T)$) increases sharply. On the other hand, as T approaches 10, k_1 remains unchanged because it is unrelated to the annuity liability (and the annuity's deferral period T). Also, k_2 remains unchanged, because it is related to $X(x_a - x_0 + T)$ through the *product* of $\partial L_{t_b}/\partial \eta^{(4)}_{t_b-x_a+1}$ and $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1}$, but $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1} = 0$ when $x_1 = 65$ and $T_1 = 11$ (because the cohort to which the q-forward is linked is already realized at $t_b = 2011$). Overall, as T approaches 10, the denominator of expression (5.41) increases sharply but the numerator remains unchanged, rendering a downward spike in the optimized hedge effectiveness.

- $x_1 = 65$ and $T_1 = 16$ (the middle panel)

  At $T = 10$, we have not only $X(x_a - x_0 + T) = X(50 - 60 + 10) = X(0)$ but also $X(T_1 - x_1 + x_a - 1) = X(16 - 65 + 50 - 1) = X(0)$. Hence, as T approaches 10, both k_2 (which is positively related to the product of $X(x_a - x_0 + T)$ and $X(T_1 - x_1 + x_a - 1)$)

[12] A larger value of $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1}$ means that $H_{t_b}(1,t_b)$ is more sensitive to $\eta^{(4)}_{t_b-x_a+1}$. As such, $H_{t_b}(1,t_b)$ is subject to more cohort effect uncertainty, and therefore k_1 – which represents the (approximate) variance $H_{t_b}(1,t_b)$ – would increase if $\partial H_{t_b}(1,t_b)/\partial \eta^{(4)}_{t_b-x_a+1}$ is larger. Similar arguments can be used to explain the other two relationships.

and k_3 (which is positively related to $X(x_a - x_0 + T)$) increase sharply. Overall, as T approaches 10, both the numerator and denominator of expression (5.41) increase significantly, resulting in an optimized hedge effectiveness that is much closer to 1.

- $x_1 = 65$ and $T_1 = 21$ (the right panel)

  The situation in this case is the same as that when $x_1 = 65$ and $T_1 = 11$, except that $\partial H_{t_b}(1, t_b)/\partial \eta^{(4)}_{t_b - x_a + 1}$ is non-zero. Still, $\partial H_{t_b}(1, t_b)/\partial \eta^{(4)}_{t_b - x_a + 1}$ is close to zero (0.0137), because it is proportional to $X(T_1 - x_1 + x_a - 1) = X(21 - 65 + 50 - 1) = X(5)$, which is very close to zero given our estimates of $\theta^{(4)}_1$ and $\phi^{(4)}_1$. Therefore, although k_2 increases as T approaches 10, the increase in k_2 tends to be mild in comparison to that in k_3. Overall, as T approaches 10, the denominator of expression (5.41) increases sharply while the numerator only increases gently, leading to a downward spike in the optimized hedge effectiveness.

One final interesting observation is that when $x_1 = 65$ and $T_1 = 16$ (i.e., when the q-forward is linked to year-of-birth 1962), the gap between the green and red lines is significant for the entire range of T except $T = 10$ (i.e., when the annuity liability is also linked to year-of-birth 1962). This observation suggests that the benefit of incorporating cohort effect uncertainty is minimal when the annuity liability and the q-forward are linked to the same cohort. To explain why, we consider the following two facts.

First, as shown in Table 5.1, $Q^*(1, 1)$ (the variance of $\eta^{(1)}_t$) and $Q^*(4, 4)$ ($(\sigma^{(4)})^2$, the variance of $\eta^{(4)}_{t-x}$) are significantly larger than the other elements in Q^*. It follows that

$$k_1 \approx \left(\frac{\partial H_{t_b}(1, t_b)}{\partial \eta^{(1)}_{t_b+1}} \right)^2 Q^*(1, 1) + \left(\frac{\partial H_{t_b}(1, t_b)}{\partial \eta^{(4)}_{t_b - x_a + 1}} \right)^2 (\sigma^{(4)})^2$$

and

$$k_2 \approx \left(\frac{\partial H_{t_b}(1, t_b)}{\partial \eta^{(1)}_{t_b+1}} \right) \left(\frac{\partial L_{t_b}}{\partial \eta^{(1)}_{t_b+1}} \right) Q^*(1, 1) + \left(\frac{\partial H_{t_b}(1, t_b)}{\partial \eta^{(4)}_{t_b - x_a + 1}} \right) \left(\frac{\partial L_{t_b}}{\partial \eta^{(4)}_{t_b - x_a + 1}} \right) (\sigma^{(4)})^2.$$

Second, when the annuity liability and the q-forward are associated with the same year-of-birth, we have $t_b + T_1 - x_1 = t_b - (x_0 - T) + 1$, which implies $T_1 - x_1 + x_a - 1 = x_a - x_0 + T = C_{t_b,1}$ and hence

$$X(T_1 - x_1 + x_a - 1) = X(C_{t_b,1}).$$

It then follows from equations (5.17), (5.18), (5.20) and (5.21) that

$$\frac{\partial H_{t_b}(1,t_b)}{\partial \eta^{(1)}_{t_b+1}} = \frac{\partial H_{t_b}(1,t_b)}{\partial \eta^{(4)}_{t_b-x_a+1}} \quad \text{and} \quad \frac{\partial L_{t_b}}{\partial \eta^{(1)}_{t_b+1}} = \frac{\partial L_{t_b}}{\partial \eta^{(4)}_{t_b-x_a+1}}.$$

Using these two facts and equations (5.39) and (5.40), we can deduce that if the annuity liability and the q-forward are linked to the same cohort, then c_1/k_1 would be close to c_2/k_2 and consequently $\hat{\mathcal{N}}_1(t_b)$ would be close to $\hat{N}_1(t_b)$ (even though c_1/k_1 and c_2/k_2 may be significant greater than zero).[13] Therefore, in this case, the sub-optimality due to the ignorance of cohort effect uncertainty is minimal.

5.6.2 The Impact of the Persistency in the Cohort Effects

The baseline results reveal that the benefit from incorporating cohort effect uncertainty depends critically on $X(s)$, which in turn depends on parameters $\theta^{(4)}_1$ and $\phi^{(4)}_1$. As $X(s)$ represents the coefficient of the lag-s innovation in infinite MA form of the cohort effect process (equation (5.12)), it may be regarded as a measure of the persistency in the cohort effects. In particular, if $X(s)$ decays rapidly as s increases from zero, then an innovation in a certain year-of-birth would have only a small impact on the cohort effect a few years-of-birth ahead; and the opposite is true if the decay in $X(s)$ is slow. On the basis of our parameter estimates, we have $\theta^{(4)}_1 + \phi^{(4)}_1 = 0.1524$, which yields a high rate of decay in $X(s)$. The baseline results are thus based on cohort effects with a very low persistency.

In this sub-section, we study how the hedging results may change if cohort effects are more persistence. To achieve this goal, we consider four hypothetical combinations of parameters $\theta^{(4)}_1$ and $\phi^{(4)}_1$, which are shown in Table 5.3. As demonstrated in Figure 5.7, the alternative parameters imply higher cohort effect persistency compared to the baseline (estimated) parameters. Under the first three sets of alternative parameters, the cohort effect process is stationary. However, under the last set of alternative parameters, the cohort effect process is a non-stationary random walk, which means that every innovation has a permanent impact on the forthcoming cohort effects. Except the assumptions about $\theta^{(4)}_1$ and $\phi^{(4)}_1$, all assumptions in the previous sub-section remain the unchanged.

The hedging results on the basis of the alternative parameter sets are shown in Figures 5.8 (when the objective is to minimize variance) and 5.9 (when the objective is to minimize Value-at-Risk). Let us first focus on the results generated using Parameter Set 1 (panels (a), (b) and (c)). One striking difference from the baseline results is that the spikes in

[13] For $T = 10$, $x_1 = 65$ and $T_1 = 16$, we have $c_1/k_1 = 0.3336$ and $c_2/k_2 = 0.2995$ (see Table 5.2).

Table 5.3: The four sets of hypothetical ARMA(1,1) parameters for the analysis in Section 5.6.2.

	$\phi_1^{(4)}$	$\theta_1^{(4)}$	$\phi_1^{(4)}+\theta_1^{(4)}$
Parameter Set 1	0.9	−0.1	0.8
Parameter Set 2	0.9	0.05	0.95
Parameter Set 3	0.9	0.3	1.2
Parameter Set 4	1	0	1

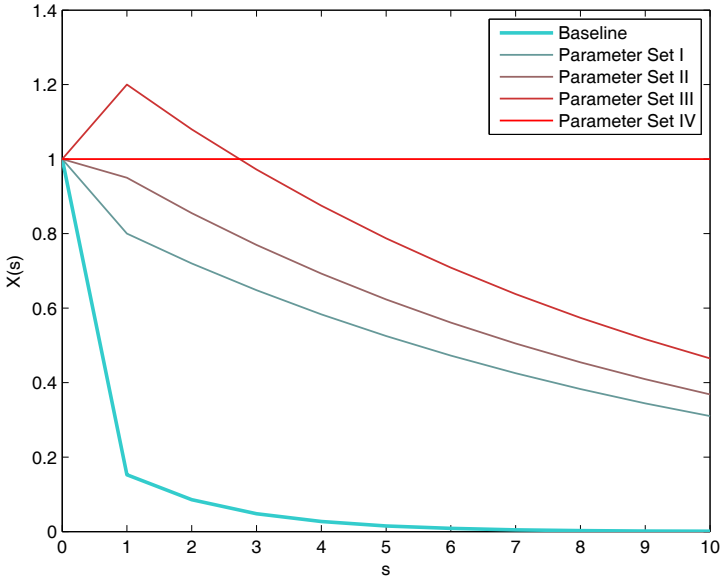


Figure 5.7: The patterns of $X(s)$ implied by the baseline (estimated) parameters and the four sets of alternative parameters. Note: $X(s) = 0$ for $s < 0$ regardless of the parameter choice.

the hedge effectiveness are less sharp. As explained in the previous sub-section, the spikes arise because $X(s)$ is particularly large at $s = 0$. Under the alternative parameter set, $X(s)$ reduces more gently as s increases from zero, and therefore sharpness of the spikes is reduced. Another striking difference is that when the alternative parameter set is used, the spike in the hedge effectiveness for $x_1 = 65$ and $T_1 = 21$ points upwards instead of downwards. As discussed in the previous sub-section, the direction of the spike depends on the magnitude of $X(5)$: the spike would point upwards if $X(5)$ is large, but downwards if $X(5)$ is zero or close to zero. For $x_1 = 65$ and $T_1 = 21$, $X(5)$ is close to zero (0.0137) under the baseline parameters, but is much higher (0.5249) under the alternative parameter set. The increase in $X(5)$ explains the change in the direction of the spike.

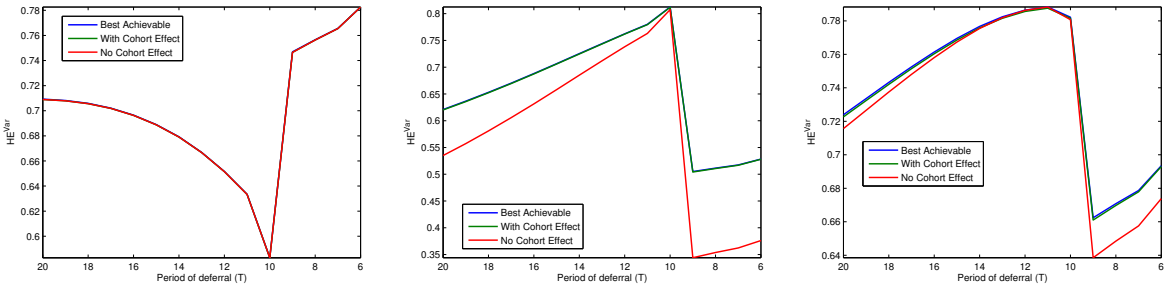
Next, we move to the results generated using Parameter Sets 2 (panels (d), (e) and (f)) and 3 (panels (g), (h) and (i)). These results are very similar to those produced using Parameter Set 1. However, because the rates of change in $X(s)$ implied by Parameter Sets 2 and 3 are even gentler, the sharpness of the spikes in the hedge effectiveness reduces even further. Also, we observe that for Parameter Set 3, the spikes occur at $T = 11$ rather than $T = 10$. The change in the location of the spikes is because $X(s)$ under Parameter Set 3 attains its maximum at $s = 1$ instead of $s = 0$.

Finally, we study the results produced using Parameter Set 4 (panels (j), (k) and (l)). Under Parameter Set 4, we have $X(s) = 1$ for all $s \geq 0$. As such, the spikes in the hedge effectiveness disappear. When a q-forward linked to a realized cohort is used (panel (j)), the hedge effectiveness is reasonably high when $T < 10$ (i.e., when the liability is also linked to a realized cohort) but becomes substantially lower when $T \geq 10$ as the q-forward is unable to mitigate the cohort effect uncertainty in the liability. When a q-forward linked to an unrealized cohort is used (panels (k) and (l)), the hedging results for $T < 10$ and $T \geq 10$ are also very different. When $T < 10$, $c_1 > 0$ but $c_2 = 0$ because the annuity liability is linked to a realized cohort. Therefore, for this range of T, $\hat{\mathcal{N}}_1(t_b)$ is quite different from $\hat{N}_1(t_b)$ and consequently the gaps between the green and red lines are significant. When $T \geq 10$, we have $X(T_1 - x_1 + x_a - 1) = X(x_a - x_0 + T) = 1$, which, according to the arguments presented in the previous sub-section, leads to $c_1/k_1 \approx c_2/k_2$ and consequently $\hat{\mathcal{N}}_1(t_b) \approx \hat{N}_1(t_b)$. Therefore, for this range of T, the gaps between the green and red lines are very narrow. It is also noteworthy that panels (k) and (l) look identical. As discussed in the previous sub-section, the impact of T_1 on the hedging results lies in the value of $X(T_1 - x_1 + x_a - 1) = X(T_1 - 65 + 50 - 1)$. Under Parameter Set 4, we have $X(T_1 - 65 + 50 - 1) = 1$ for any $T_1 \geq 16$, and therefore the results shown in panels (k) and (l) are the same.

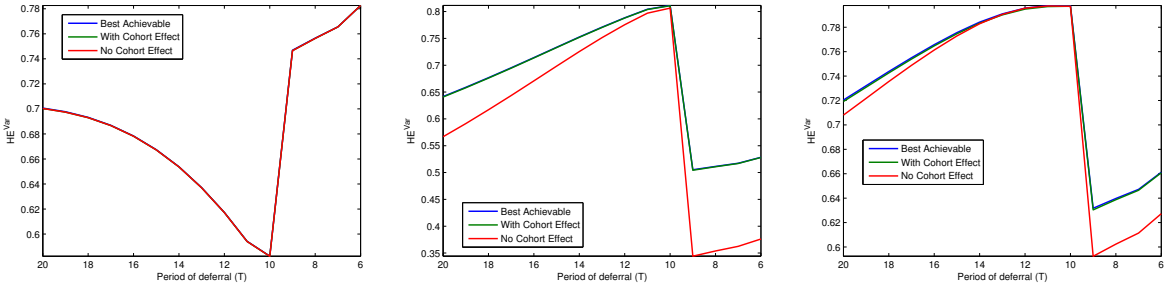
5.6.3 The Effects of τ and λ

The baseline results assume that $\tau = 1$ and $\lambda = 0$. For completeness, in this section we sensitivity test τ and λ, while keeping all other assumptions unchanged.

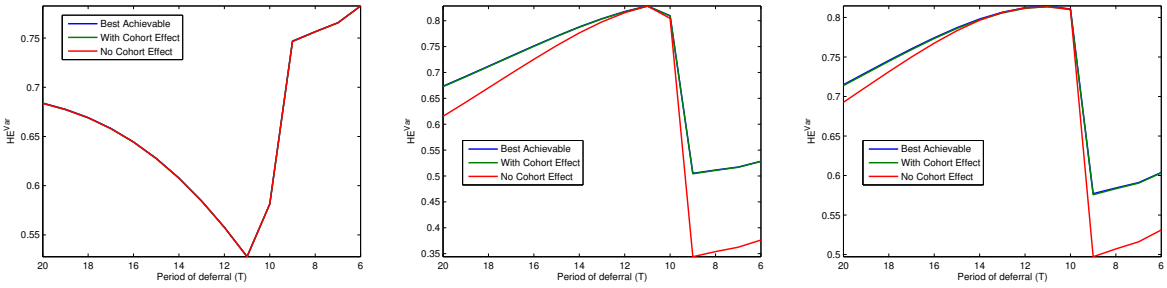
Figure 5.10 shows the relationship between the calculated hedge effectiveness and the hedging horizon τ. The results are based on a deferral period of $T = 20$ years and a q-forward with a reference age of $x_1 = 65$ and a time-to-maturity of $T_1 = 16$ years. It can be observed that the calculated hedge effectiveness increases with τ when $\tau \leq 16$, but decreases when $\tau > 16$. We may understand this observed relationship by noting that the optimized value of HE^{Var} is $k_2^2/(k_1 k_3)$ and by revisiting the expressions for k_1, k_2 and k_3 (equations (5.26) to (5.28)). For $\tau \leq T_1$, k_1, k_2 and k_3 all increase monotonically with τ, as



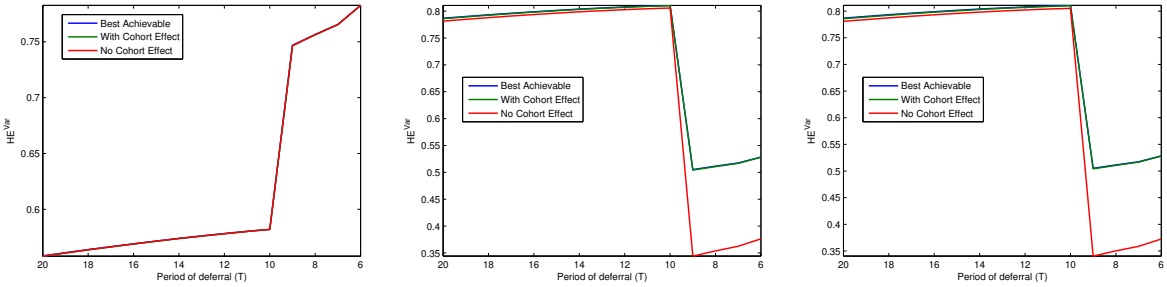
(a) Set 1; $x_1 = 65$, $T_1 = 11$ (b) Set 1; $x_1 = 65$, $T_1 = 16$ (c) Set 1; $x_1 = 65$, $T_1 = 21$



(d) Set 2; $x_1 = 65$, $T_1 = 11$ (e) Set 2; $x_1 = 65$, $T_1 = 16$ (f) Set 2; $x_1 = 65$, $T_1 = 21$

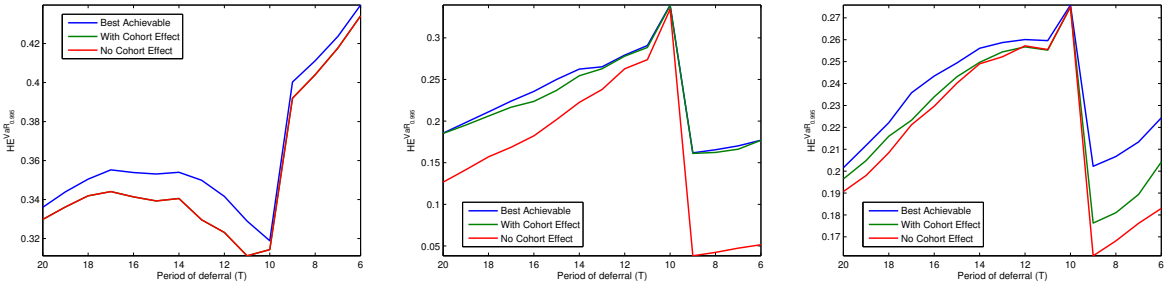


(g) Set 3; $x_1 = 65$, $T_1 = 11$ (h) Set 3; $x_1 = 65$, $T_1 = 16$ (i) Set 3; $x_1 = 65$, $T_1 = 21$

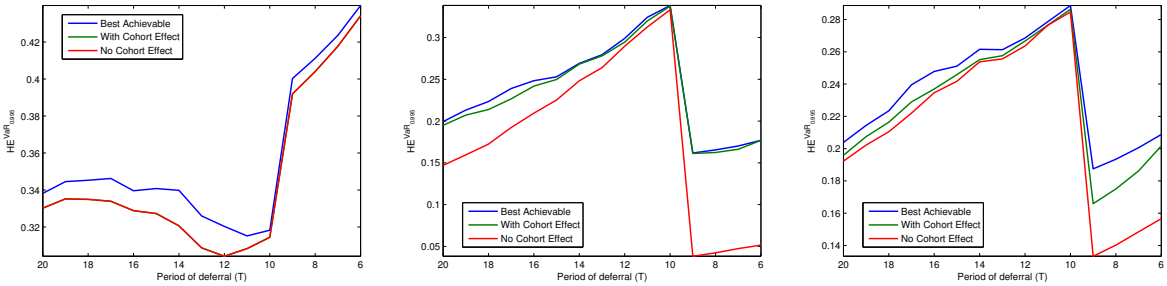


(j) Set 4; $x_1 = 65$, $T_1 = 11$ (k) Set 4; $x_1 = 65$, $T_1 = 16$ (l) Set 4; $x_1 = 65$, $T_1 = 21$

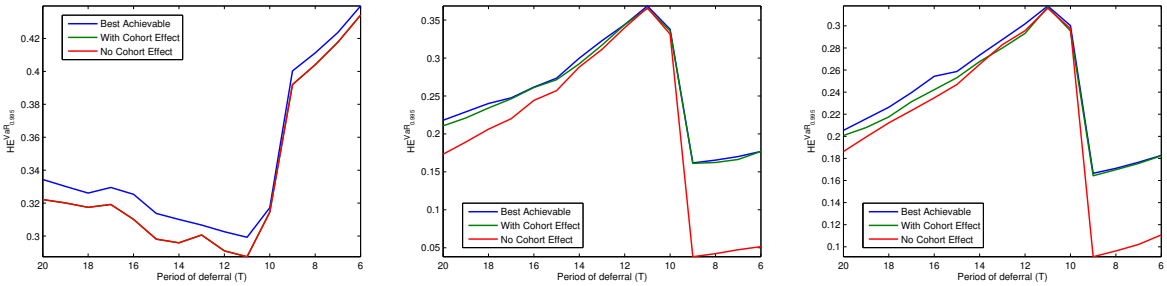
Figure 5.8: The calculated values of HE^{Var} under the four parameter sets specified in Table 5.3.



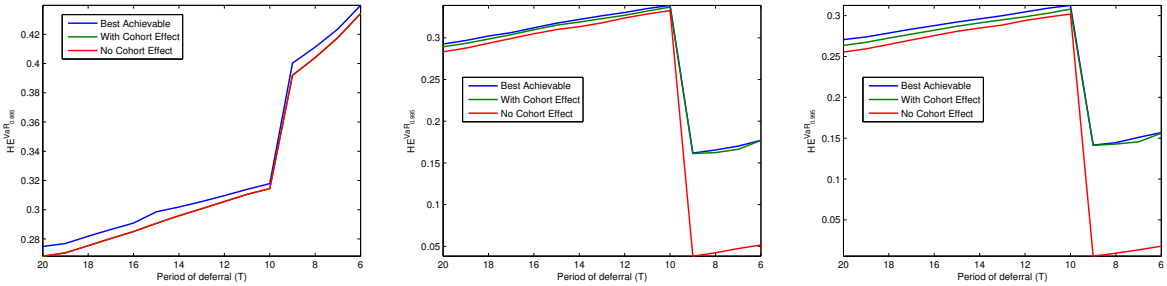
(a) Set 1: $x_1 = 65$, $T_1 = 11$ (b) Set 1: $x_1 = 65$, $T_1 = 16$ (c) Set 1: $x_1 = 65$, $T_1 = 21$



(d) Set 2: $x_1 = 65$, $T_1 = 11$ (e) Set 2: $x_1 = 65$, $T_1 = 16$ (f) Set 2: $x_1 = 65$, $T_1 = 21$

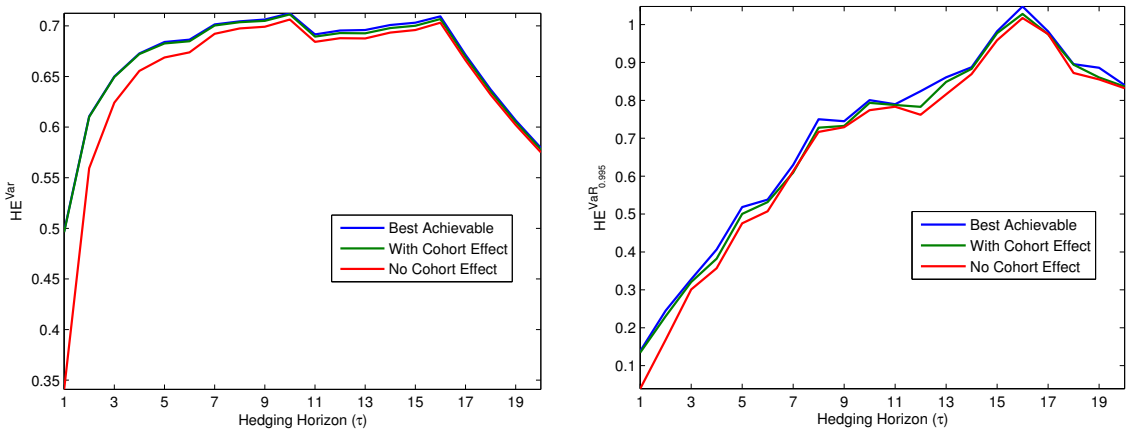


(g) Set 3: $x_1 = 65$, $T_1 = 11$ (h) Set 3: $x_1 = 65$, $T_1 = 16$ (i) Set 3: $x_1 = 65$, $T_1 = 21$



(j) Set 4: $x_1 = 65$, $T_1 = 11$ (k) Set 4: $x_1 = 65$, $T_1 = 16$ (l) Set 4: $x_1 = 65$, $T_1 = 21$

Figure 5.9: The calculated values of $\mathrm{HE}^{\mathrm{VaR}_{0.995}}$ under the four parameter sets specified in Table 5.3.



(a) Variance Minimization

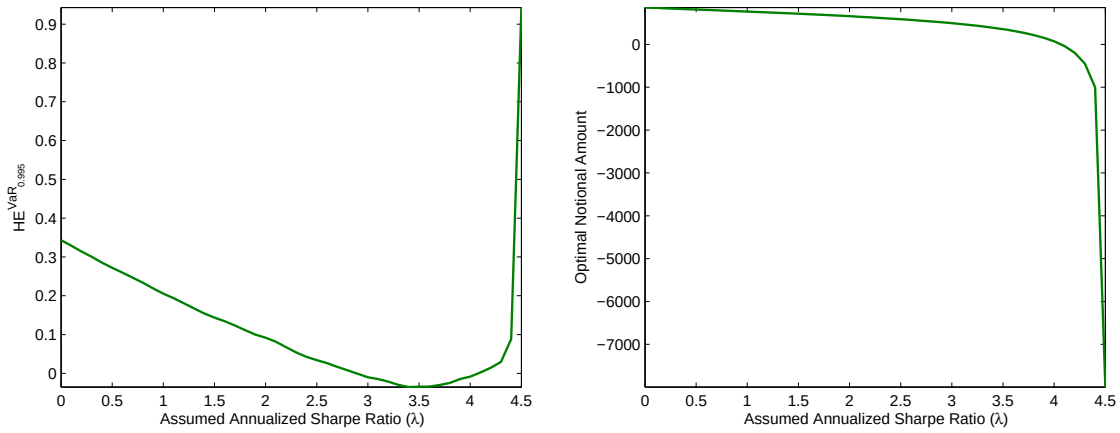
(b) Value-at-Risk Minimization

Figure 5.10: The calculated values of HE^{Var} and $\text{HE}^{\text{VaR}_{0.995}}$ for different hedging horizons. It is assumed that $T = 20$, $x_1 = 65$ and $T_1 = 16$.

every term in the summands in equations (5.26) to (5.28) is non-negative. In this example, k_2^2 increases at a faster rate than $k_1 k_3$, leading to a rise in the optimized HE^{Var} for $\tau \leq 16$. However, for $\tau > T_1$, k_2 may reduce with τ, because $\partial H_{t_b}(1, t_b)/\partial \vec{\eta}_{t_b+s} = 0$ for $s > T_1$ (as the value of the q-forward is unrelated to any innovation vector beyond the q-forward's maturity) and $\partial L_{t_b}/\partial \vec{\eta}_{t_b+s} < 0$ (as the value of the liability increases when the values of the innovations decrease). The possible reduction in k_2 explains the observed downward trend when $\tau > 16$.

We also observe in Figure 5.10 that the benefit of incorporating cohort effect uncertainty diminishes as τ increases. This observation follows from the specifications of the annuity liability and the q-forward. In this example, both the annuity liability and the q-forward are linked to year-of-birth 1962 (one year ahead of the last year-of-birth covered by the data sample). As such, no matter how large τ is, the only unrealized cohort-effect-related innovation involved in $V_{L_{t_b}}(t_b + \tau)$ and $V_{H_{t_b}(1,t_b)}(t_b + \tau)$ is the one that is one step ahead of the forecast origin. On the other hand, as τ increases, the number of unrealized period-effect-related innovations involved in $V_{L_{t_b}}(t_b + \tau)$ and $V_{H_{t_b}(1,t_b)}(t_b + \tau)$ increases.[14] Overall, when τ becomes higher, cohort effect uncertainty plays a less significant role, and hence the benefit of incorporating cohort effect uncertainty reduces.

[14]It follows from the definitions in Section 5.3 that $V_{L_{t_b}}(t_b+\tau)$ involves the period-effect-related innovations from time t_b+1 to $t_b+\min(\tau, T+\omega-x_0)$, whereas $V_{H_{t_b}(1,t_b)}(t_b+\tau)$ involves to the period-effect-related innovations from time $t+1$ to $t+\min(\tau, T_j)$.



(a) $\text{HE}^{\text{VaR}_{0.995}}$ (b) Optimized notional amount

Figure 5.11: The calculated values of $\text{HE}^{\text{VaR}_{0.995}}$ and the corresponding optimized notional amounts for $0 \leq \lambda \leq 4.5$. It is assumed that $T = 20$, $x_1 = 65$ and $T_1 = 16$.

As explained in Section 5.4, if the objective is to minimize variance, then the market price of risk λ demanded by the counterparty (the fixed-rate payer) has no impact on the optimal notional amounts. However, if the objective is to minimize Value-at-Risk, then the value of λ affects the optimal notional amounts and consequently the hedge effectiveness. Figure 5.11 shows the calculated values of $\text{HE}^{\text{VaR}_{0.995}}$ and the corresponding optimized notional amounts for different assumed values of λ. The results are based again on the assumptions that $T = 20$, $x_1 = 65$ and $T_1 = 16$ years, and have incorporated cohort effect uncertainty.

When λ is greater than zero, the q-forward is no longer costless. As the hedge cost (which is proportional to the notional amount) offsets the benefit from the correlation effects, it is optimal to acquire a smaller notional amount and the optimal hedge effectiveness reduces accordingly. This argument explains what we observe when λ increases from 0 to 3.4. When $\lambda = 3.4$, the hedge cost completely offsets the benefit from the correlation effects, leading both the optimal notional amount and the hedge effectiveness to zero. Mathematically, equation (5.34) implies that $(\hat{H}_{t_b}(1, t_b))^2 = 0$ when $\lambda = 0$ and that $(\hat{H}_{t_b}(1, t_b))^2$ increases with λ. At $\lambda = 3.4$, $(\hat{H}_{t_b}(1, t_b))^2$ reaches $(\Phi^{-1}(0.995))^2(k_2)^2/k_3$, resulting in $c = 0$ and thus $\hat{N}_1(t_b) = 0$.

A higher λ means a more generous reward to the fixed-rate payer. When λ increases further, the reward to the fixed-rate payer becomes so high that the annuity provider would be better off if it participates in the q-forward as a fixed-rate payer rather than a

fixed-rate receiver. For this reason, when λ is larger than 3.4, the optimal notional amount becomes negative (which means a participation as a fixed-rate payer). As λ increases even further, the reward to the fixed-rate payer would eventually reach a level that a more negative notional amount would always bring more benefit to the annuity provider. In this case, the optimal notional amount does not exist. Mathematically, when λ exceeds 4.5, $(\hat{H}_{t_b}(1,t_b))^2$ is strictly greater than $(\Phi^{-1}(0.995))^2 k_1$,[15] violating condition (5.33) and hence resulting in no solution to $\hat{N}_1(t_b)$.

5.7 Conclusion

Cohort effect uncertainty often plays a role in an index-based longevity hedge, because either the liability being hedged, the hedging instrument, or both are associated with cohorts that are not yet realized. However, it is largely overlooked in the existing methods for calibrating index-based longevity hedges. When cohort effect uncertainty is present but is not taken into account in the hedging strategy, the resulting longevity hedge may perform sub-optimally.

In this chapter, we have contributed a hedge calibration method that fully incorporates cohort effect uncertainty. With the proposed method, one can find the notional amounts that minimizes the variance or Value-at-Risk of the τ-year ahead values of the hedged position. The optimized notional amounts can be expressed in terms of Q^* (the covariance matrix of the innovation vector) and various partial derivatives that can be analytically computed. By setting $\tau = 1$ and executing the calibration method every year, an annually-adjusted dynamic longevity hedge can be created easily.

We have also contributed a method that eases the computation burden required in evaluating a value longevity hedge. Rather than simulating sample paths of future period and cohort effects, the proposed method evaluates a longevity hedge by simulating sample paths of future innovations. As all innovations are serially uncorrelated by definition, we can always use the same collection of sample paths of innovations for valuing the hedge portfolio at the end of the hedging horizon. Nested simulations can therefore be avoided.

The proposed methods have been illustrated using historical mortality data from English and Welsh male population, for which cohort effects are found to be significant. Through an illustrative longevity hedge with one q-forward, we have examined the benefit of incorporating cohort effects under different circumstances. A number of lessons can be learnt from the empirical work.

[15] It is argued in Section 5.4.4 that $0 < k_2/(k_1 k_3) < 1$ and $k_1 > 0$. Thus, $(\Phi^{-1}(0.995))^2 k_1$ must be strictly greater than $(\Phi^{-1}(0.995))^2 (k_2)^2/k_3$.

First, the benefit of incorporating cohort effects depends heavily on the year-of-birth to which the q-forward is linked. For a static hedge established at time t_b, the benefit tends to be more significant if the difference between the year-of-birth associated with the q-forward $(t_b + T_1 - x_1)$ and the last year-of-birth covered by the data sample $(t_b - x_a)$ is small. The reason behind is that when $(t_b + T_1 - x_1) - (t_b - x_a)$ is close to zero, cohort effects are influential in the variance of the q-forward's values and the covariance between the values of the q-forward and the liability being hedged.

Second, even if the liability being hedged is not linked to any unrealized cohort, incorporating cohort effect uncertainty into the calibration process may still yield a more effective hedge. This outcome is because the optimal notional amount depends on the sensitivity of the q-forward's value to the underlying cohort effect. As long as the q-forward is linked to an unrealized cohort, ignoring this component would result in a sub-optimal notional amount.

Third, the benefit of incorporating cohort effect uncertainty is small when the liability being hedged and the q-forward are linked to the same cohort, because under this condition c_1/k_1 and c_2/k_2 are close to each other. This finding is complementary to a conclusion obtained by Cairns et al. (2014): a longevity hedge tends to be less effective when there is a mismatch between the cohorts to which the hedging instrument and the liability being hedged are linked. When the problem of mismatched cohorts is inevitable, the hedger should be mindful of the less than ideal hedge effectiveness and factor cohort effect uncertainty into his/her hedging strategy.

Fourth, the hedging results depend heavily on $X(s)$, which measures the persistency of cohort effects. For instance, if cohort effects are more persistent, then the hedging results would be more robust relative to the liability's period of deferral. Also, if cohort effects are more persistent, then the benefit of incorporating cohort effect uncertainty would be less sensitive to the specification of the q-forward used.

Fifth and last, the benefit of incorporating cohort effect uncertainty diminishes as the hedging horizon τ increases. The reason behind is that in $V_{L_{t_b}}(t_b + \tau)$ and $V_{H_{t_b}(1,t_b)}(t_b + \tau)$, the proportion of the innovations that are related to cohort effects falls as τ rises. Because cohort effect uncertainty tends to be more significant when τ is small, it deserves some attention when the hedger's objective is to, for example, mitigate the one-year risk defined in Solvency II.

As in many studies on longevity risk modeling, we assume in this chapter that the innovations in the period and cohort effect processes are normally distributed. Recently, it has been found that such innovations could be leptokurtic, thereby calling for a non-normal distributional assumption (see, e.g., Wang et al., 2015). When a different distributional

assumption is used, the theoretical result for variance minimization would remain the same but that for Value-at-Risk minimization would change. One possible direction for further research is to study how the strategy for minimizing Value-at-Risk should be adapted if the innovations follow a more sophisticated distribution.

To focus on cohort effect uncertainty, we do not consider population basis risk in this chapter. However, with the aid of a multi-population mortality model, we can easily extend the proposed calibration method to incorporate the difference between the populations involved. For example, if we assume the augmented common factor model (Li and Lee, 2005), then we can derive the hedging strategy by first formulating L_t and $H_t(j,t)$ as functions of the innovations from the random walk that captures the common mortality trend and the autoregressive processes that characterize the population-specific mortality trends. Another caveat of this chapter is that sampling risk is ignored. To gauge the impact of sampling risk, one may use a discrete distribution such as Poisson to model death counts conditioned on the realized underlying mortality rate (see, e.g., Cairn et al., 2014). We leave these extensions to future research.

Chapter 6

Concluding Remarks

Longevity risk is comprised of various risk factors. In this thesis, we have explored four of the most important risk factors constituting longevity risk: mortality jump risk, longevity drift risk, population basis risk and cohort mismatch risk.

Mortality jump risk refers to the uncertainty associated with catastrophic mortality events. In previous studies, most focus has been placed on modeling the severity and frequency of mortality jumps. One characteristic that is important but often ignored is the age pattern of mortality jumps. In this thesis, we have investigated how the mortality jump effect can be distributed across different ages. We have shown how this age pattern can be modeled explicitly by introducing two variants of the Lee-Carter model. The proposed model variants use a distinct collection of parameters to capture the age pattern of mortality jumps. The key advantage of our modeling approach is that it permits a random age pattern of mortality jump effects.

Since mortality jumps are infrequent, we have addressed parameter risk. We have conducted a parametric bootstrapping procedure to quantify the uncertainty surrounding each of the model parameters. We found that the standard errors of some jump-effect-related parameters are quite high, due to the fact that only a handful of extreme mortality events occurred during the data sample period. We have also shown that the impact of parameter uncertainty on pricing depends quite heavily on the age range with which the security being priced is associated.

Different from mortality jump risk which only has a short-term effect, longevity drift risk is the uncertainty associated with the longevity trend. Longevity drift risk affects the mortality dynamics over a long time. Most of the existing mortality models focus on capturing the uncertainty surrounding a specific trend, leaving the uncertainty related

to the trend itself unmodeled. To incorporate longevity drift risk, we have introduced the LLCBD model which is a locally-linear extension to the original CBD model. The LLCBD model is shown to be demographically intuitive and can be estimated in one single stage using EM algorithm and Kalman filter. Compared to the original CBD model, the LLCBD model provides better goodness-of-fit in terms of AIC, yields more accurate short- and long-term forecasts in terms of mean error and mean squared error, generates forecasts that are more consistent with the observed trends in the recent past and are more robust relative to changes in the length of the calibration window.

We have also developed a new hedging method called the generalized state-space hedging method from which one can construct an index-based longevity hedge to mitigate drift risk. The proposed hedging method can ameliorate the problems of sub-optimality and singularity that the traditional delta and delta-nuga hedging methods are subject to. We have further applied the generalized state-space hedging method to a hypothetical scenario and shown the importance of managing drift risk.

Population basis risk, which arises from population differences, is inevitable in an index-based longevity hedge. To measure and manage population basis risk, the proposed generalized state-space hedging method is then extended to a multi-population setting. The extended hedging method allows the calibration of a longevity hedge that is composed of standardized hedging instruments which take into account the difference in mortality improvements between the populations associated with the hedging instruments and the hedger's portfolio. The extension may be regarded as a general framework, because it is applicable to all coherent multi-population stochastic mortality models that can be written in state-space representations.

In addition, based on the proposed hedging method, we have also investigated how we can analytically approximate the variances of an unhedged position. This approximation method enables the calculation of hedge effectiveness without using simulation. Using this approximation method, we can derive an empirical relationship between hedge effectiveness and the composition of the hedge portfolio and then analyse longevity risk through a variance decomposition. A new metric called standardized basis risk profile has been proposed which allows us to assess the relative levels of population basis risk that q-forwards with different reference populations, reference ages, and times-to-maturity may lead to.

Other than population basis risk, cohort mismatch risk also plays a role in an index-based longevity hedge. When cohort effect uncertainty is present but not considered in the hedging strategy, the resulting longevity hedge may also perform sub-optimally. To incorporate cohort mismatch risk, we have developed a new hedge calibration method through which one can analytically obtain the optimal notional amounts that minimize the

variance or Value-at-Risk of the τ-year ahead values of the hedged position. As well, we have designed a method that eases the computation burden required in evaluating a value longevity hedge. The proposed method evaluates a longevity hedge using the simulated sample paths of future innovations. As all innovations are serially uncorrelated, we can always use the same collection of innovations for valuing the hedge portfolio at the end of the hedging horizon and therefore avoid nested simulations.

Using historical mortality data from English and Welsh male population, we have examined the benefit of incorporating cohort effects in a longevity hedge under different circumstances. From the empirical work, a number of lessons can be learned regarding the incorporation of cohort mismatch risk:

- the benefit of incorporating cohort effects depends heavily on the year-of-birth to which the q-forward is linked;
- even if the liability being hedged is not linked to any unrealized cohort, incorporating cohort effect uncertainty into the calibration process may still yield a more effective hedge;
- the benefit of incorporating cohort effect uncertainty is small when the liability being hedged and the q-forward are linked to the same cohort;
- the hedging results depend heavily on the persistency of cohort effects; and
- the benefit of incorporating cohort effect uncertainty diminishes as the hedging horizon increases.

For future research, one may be interested in generalizing model J1, model J2 (proposed in Chapter 2), and the LLCBD model (proposed in Chapter 3) to a multi-population version, which would be useful for pricing and analysing longevity products that are linked to a group of populations. One may also use the modeling approach described in Chapter 2 to capture the age patterns of intermittent changes in long-term mortality trends.

We acknowledge that cohort effects are significant in certain populations, and that it is not trivial to incorporate cohort effects in a state-space representation in which the vector of hidden states evolve over time rather than year of birth. Therefore, it would also be interesting to investigate how the LLCBD model and the generalized state-space hedging method can be further extended to incorporate cohort effects.

As in many studies on longevity risk modeling, we assume in this thesis that the innovations in the period and cohort effect processes are normally distributed. Recently, it has

been found that such innovations could be leptokurtic, thereby calling for a non-normal distributional assumption (see, e.g., Wang et al., 2015). When a different distributional assumption is used, the theoretical result for variance minimization would remain the same but that for Value-at-Risk minimization would change. Another possible direction for further research is to study how the strategy for minimizing Value-at-Risk should be adapted if the innovations follow a more sophisticated distribution.

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, **AC-19**, 716-723.

Ahmadi, S. and Li, J.S.-H. (2014). Coherent Mortality Forecasting with Generalized Linear Models: A Modified Time-Transformation Approach. *Insurance: Mathematics and Economics*, **59**, 194-221.

Bauer, D. and Kramer, F.W. (2009). Risk and Valuation of Mortality Contingent Catastrophe Bonds. Working Paper, Georgia State University.

Biffis, E. (2005). Affine Process for Dynamic Mortality and Actuarial Valuations. *Insurance: Mathematics and Economics*, **37**, 443-468.

Blake, D., Cairns, A., Coughlan, G., Dowd, K. and MacMinn, R. (2013). The New Life Market. *Journal of Risk and Insurance*, **80**, 501-557.

Blake, D., Cairns, A.J.G. and Dowd, K. (2006). Living with Mortality: Longevity Bonds and other Mortality-Linked Securities. *British Actuarial Journal*, **12**, 153-197.

Blake, D., MacMinn, R., Li, J.S.-H. and Hardy, M. (2014). Longevity Risk and Capital Markets: The 2012-2013 Update. *North American Actuarial Journal*, **18**, 501-557.

Booth, H., Maindonald, J., and Smith, L. (2002). Applying Lee-Carter under Conditions of Variable Mortality Decline. *Population Studies*, **56**, 325-336.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control.* 2nd edition. San Francisco: Holden-Day.

Brouhns, N., Denuit, M. and Van Keilegom, I. (2005). Bootstrapping the Poisson Log-Bilinear Model for Mortality Forecasting. *Scandinavian Actuarial Journal*, **2005**, 212-224.

Cairns, A.J.G. (2011). Modelling and Management of Longevity Risk: Approximations to Survival Functions and Dynamic Hedging. *Insurance: Mathematics and Economics*, **49**, 438-453.

Cairns, A.J.G. (2013). Robust Hedging of Longevity Risk. *Journal of Risk and Insurance*, **80**, 621-648.

Cairns, A.J.G., Blake, D., and Dowd, K., (2006) A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, **73**, 687-718.

Cairns, A.J.G., Blake, D., Dowd, K., and Coughlan, G.D. (2014). Longevity Hedge Effectiveness: A Decomposition. *Quantitative Finance*, **14**, 217-235.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011a). Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics and Economics*, **48**, 355-367.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A Quantitative Comparison of Stochastic Mortality Models using Data from England and Wales and the United States. *North American Actuarial Journal*, **13**, 1-35.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D. and Khalaf-Allah, M. (2011b). Bayesian Stochastic Mortality Modelling for Two Populations. *ASTIN Bulletin*, **41**, 29-59.

Canadian Institute of Actuaries (2014). Final Report on Canadian Pensioners' Mortality. Available at http://www.cia-ica.ca/docs/default-source/2014/214013e.pdf.

Carter, L.R. (1996). Forecasting U.S. Mortality: A Comparison of Box-Jenkins ARIMA and Structural Time Series Models. *The Sociological Quarterly*, **37**, 127-144.

Cavanaugh, J., and Shumway, R. (1997). A Bootstrap Variant of AIC for State-Space Model Selection. *Statistica Sinica*, **7**, 473-496.

Chen, H. (2013). A Family of Mortality Jump Models Applied to US Data. *Asia-Pacific Journal of Risk and Insurance*, **8**, 105-121.

Chen, H. and Cox, S.H. (2009). Modeling Mortality with Jumps: Applications to Mortality Securitization. *Journal of Risk and Insurance*, **76**, 727-751.

Chen, H., Cox, S.H. and Wang, S.S. (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. *Insurance: Mathematics and Economics*, **46**, 371-384.

Chen, H. and Cummins, J.D. (2010). Longevity Bond Premiums: The Extreme Value Approach and Risk Cubic Pricing. *Insurance: Mathematics and Economics*, **46**, 150-161.

Chen, H. MacMinn, R.D. and Sun, T. (2013a). Multi-Population Mortality Models: A Factor Copula Approach. Paper presented at the Ninth International Longevity Risk and Capital Markets Solutions Symposium, Beijing, China.

Chen, H., Sherris, M., Sun, T. and Zhu, W. (2013b). Living with Ambiguity: Pricing Mortality-Linked Securities with Smooth Ambiguity Preferences. *Journal of Risk and Insurance*, **80**, 705-732.

Chuang, S.-L. and Brockett, P.L. (2014). Modeling and Pricing Longevity Derivatives using Stochastic Mortality Rates and the Esscher Transform. *North American Actuarial Journal*, **18**, 22-37.

Coelho, E., and Nunes, L. (2011). Forecasting Mortality in the Event of a Structural Change. *Journal of the Royal Statistical Society Series A*, **174**, 713736.

Continuous Mortality Investigation Bureau (2002). An Interim Basis for Adjusting the "92" Series Mortality Projections for Cohort Effects. CMI Working Paper 1. London: Institute of Actuaries and Faculty of Actuaries.

Continuous Mortality Investigation Bureau (2009a). A Prototype Mortality Projections Model, Part One – An Outline of the Proposed Approach. Working Papers 38, available at http://www.actuaries.org.uk/research-and-resources/documents/cmi-working-paper-38-prototype-mortality-projections-model-part-one.

Continuous Mortality Investigation Bureau (2009b). A Prototype Mortality Projections Model, Part Two – Detailed Analysis. Working Papers 39, available at http://www.actuaries.org.uk/research-and-resources/documents/cmi-working-paper-39-prototype-mortality-projections-model-part-two.

Coughlan, G. (2009). Longevity Risk Transfer: Indices and Capital Market Solutions. In Barrieu, P.M. and Albertini, L. (eds.) *The Handbook of Insurance Linked Securities*. London: Wiley.

Coughlan, G., Blake, D., MacMinn, R., Cairns, A.J.G. and Dowd, K. (2013). Longevity Risk and Hedging Solutions. In *Handbook of Insurance* (pp. 997-1035). Springer: New York.

Coughlan, G.D., Khalaf-Allah, M., Ye, Y., Kumar, S., Cairns, A.J.G., Blake, D., and Dowd, K. (2011). Longevity Hedging 101: A Framework for Longevity Basis Risk Analysis and Hedge Effectiveness. *North American Actuarial Journal*, **15**, 150-176.

Coughlan, G., Epstein, D., Sinha, A. and Honig, P. (2007). q-Forwards: Derivatives for Transferring Longevity and Mortality Risk.

Cox, N.J., Tamblyn, S.E., Tam, T. (2003). Influenza Pandemic Planning. *Vaccine*, **21**, 1801-1803.

Cox, S.H., Lin, Y. and Pedersen, H. (2010). Mortality Risk Modeling: Applications to Insurance Securitization. *Insurance: Mathematics and Economics*, **46**, 242-253.

Cox, S.H., Lin, Y. and Wang, S. (2006). Multivariate Exponential Tilting and Pricing Implications for Mortality Securitization. *Journal of Risk and Insurance*, **73**, 113-136.

Crosby, A.W. (1976). Epidemic and Peace, 1918. Westford, CT: Greenwood Press.

Dahl, M. (2004). Stochastic Mortality in Life Insurance: Market Reserves and Mortality-Linked Insurance Contracts. *Insurance: Mathematics and Economics*, **35**, 113-136.

Dahl, M., Melchior, M. and Møller, T. (2008). On Systematic Mortality Risk and Risk Minimization with Mortality Swaps. *Scandinavian Actuarial Journal*, **108**, 114-146.

Dahl, M. and Møller, T. (2006). Valuation and Hedging of Life Insurance Liabilities with Systematic Mortality Risk. *Insurance: Mathematics and Economics*, **39**, 193-217.

Dauer, C.C. and Serfling, R.E. (1961). Mortality from Influenza. 1957-58 and 1959-60. *American Reviews of Respiratory Disease*, **83**, 15-28.

Davison, A.C. (2003). Statistical Models. Cambridge, United Kingdom: Cambridge University Press.

D'Amato, V., Haberman, S., Piscopo, G., Russolillo, M., and Trapani, L. (2014). Detecting Common Longevity Trends by a Multiple Population Approach. *North American Actuarial Journal*, **18**, 139-149.

de Jong, P., and Tickle, L. (2006). Extending Lee-Carter Mortality Forecasting. *Mathematical Population Studies*, **13**, 1-18.

Deng, Y., Brockett, P.L. and MacMinn, R.D. (2012). Longevity/Mortality Risk Modeling and Securities Pricing. *Journal of Risk and Insurance*, **79**, 697-721.

Denuit, M. and Dhaene, J. (2007). Comonotonic Bounds on the Survival Probability in the Lee-Carter Model for Mortality Projection. *Computational and Applied Mathematics*, **203**, 169-176.

Denuit, M., and Goderniaux, A. (2005). Closing and Projecting Lifetables Using Log-Linear Models. *Bulletin of the Swiss Association of Actuaries*, **1**, 29-49.

Denuit, M., Haberman, S. and Renshaw, A.E. (2010). Comonotonic Approximations to Quantiles of Life Annuity Conditional Expected Present Values: Extensions to General ARIMA Models and Comparison with the Bootstrap. *ASTIN Bulletin*, **40**, 331-349.

Diao, L. and Cook, R.J. (2014). Composite likelihood for joint analysis of multiple multistate processes via copulas. *Biostatistics*, **15**, 690-705.

Dowd, K., Blake, D., Cairns, A.J.G. and Coughlan, G.D. (2011a). Hedging Pension Risks with the Age-Period-Cohort Two-Population Gravity Model. In: Seventh International Longevity Risk and Capital Markets Solutions Conference, Frankfurt, September 2011.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010a). Evaluating the Goodness of Fit of Stochastic Mortality Models. *Insurance: Mathematics and Economics*, **47**, 255-265.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010b). Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts. *North American Actuarial Journal*, **14**, 281-298.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D. and Khalaf-Allah, M. (2011b). A Gravity Model of Mortality Rates for Two Related Populations. *North American Actuarial Journal*, **15**, 331-356.

Frittelli, M. (2000). The Minimal Entropy Martingale Measure and the Valuation Problem in Incomplete Market. *Mathematical Finance*, **10**, 39-52.

Gallop, A. (2006). Mortality Improvements and Evolution of Life Expectancies. Paper presented at the Seminar on Demographic, Economic and Investment Perspectives for Canada, Office of the Superintendent of Financial Institutions Canada.

Girosi, F., and King, G. (2005) A Reassessment of Lee-Carter Mortality Forecasting Method. *International Journal of Forecasting*, **21**, 249-260.

Gong, G. and Webb, A. (2010). Evaluating the Advanced Life Deferred Annuity – An Annuity People Might Actually Buy. *Insurance: Mathematics and Economics*, **46**, 210-221.

Graziani, G. (2014). Longevity risk - A Fine Balance. *Institutional Investor Journals: Special Issue on Pension and Longevity Risk Transfer for Institutional Investors*, **2014**, 35-27.

Haberman, S., Kaishev, V., Millossovich, P, Villegas, A., Baxter, S. Gaches, A., Gunnlaugsson, S. and Sison, M. (2014). Longevity Basis Risk: A Methodology for Assessing Basis Risk. Research investigation and report by Cass Business School and Hymans Robertson LLP for the Institute and Faculty of Actuaries and the Life and Longevity Markets Association. Available at

http://www.actuaries.org.uk/research-and-resources/documents/sessional-research-event-longevity-basis-risk-methodology

Haberman, S. and Renshaw, A. (2012). Parametric Mortality Improvement Rate Modelling and Projecting. *Insurance: Mathematics and Economics*, **50**, 309-333.

Hainaut, D. and Devolder, P. (2008). Mortality Modelling with Levy Processes. *Insurance: Mathematics and Economics*, **42**, 409-418.

Hári, N., de Waegenaere, A., Melenberg, B., and Nijman, T.E. (2008). Estimating the Term Structure of Mortality. *Insurance: Mathematics and Economics*, **42**, 492-504.

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge university press.

Hatzopoulos, P. and Haberman, S. (2013). Common Mortality Modeling and Coherent Forecasts. An Empirical Analysis of Worldwide Mortality Data. *Insurance: Mathematics and Economics*, **52**, 320-337.

Holmes, E.C. (2004). Enhanced: 1918 and All That. *Science*, **303**, 1787-1788.

Holmes, E.E. (2013). Derivation of an EM algorithm for Constrained and Unconstrained Multivariate Autoregressive State-Space (MARSS) Models. arXiv preprint arXiv:1302.3919.

Horneff, W., Maurer, R. and Rogalla, R. (2010). Dynamic Portfolio Choice with Deferred Annuities. *Journal of Banking and Finance*, **34**, 2652-2664.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute of Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 1 April 2015).

Hunt, A. and Blake, D. (2015). Modelling Longevity Bonds: Analysing the Swiss Re Kortis Bond. *Insurance: Mathematics and Economics*, **63**, 12-29.

Imhof, J.P. (1961). Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika*, **48**, 419-426.

Jarner, S.F. and Kryger, E.M. (2011). Modelling Adult Mortality in Small Populations: The SAINT Model. *ASTIN Bulletin*, **41**, 377-418.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Fluids Engineering*, **82**, 35-45.

Kalman, R.E., and Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Fluids Engineering*, **83**, 95-108.

Kannisto, V., Lauristen, J., Thatcher, A.R., and Vaupel, J.W. (1994). Reduction in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries. *Population Development Review*, *20*, 793-810.

Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.

LaMotte, L.R. and McWhorter, A. (1978). An Exact Test for the Presence of Random Walk Coefficients in a Linear Regression Model. *Journal of the American Statistical Association*, **364**, 816-820.

Lee, R. and Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, **87**, 659-671.

Lee, R. and Miller, T. (2001). Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality . *Demography*, **38**, 537-549.

Li, H., De Waegenaere, A., and Melenberg, B. (2015a). The Choice of Sample Size for Mortality Forecasting: A Bayesian Learning Approach. *Insurance: Mathematics and Economics*, **63**, 153-168.

Li, J.S.-H. (2010). Pricing Longevity Risk with the Parametric Bootstrap: a Maximum Entropy Approach. *Insurance: Mathematics and Economics*, **47**, 176-186.

Li, J.S.-H., Chan, W.S. and Cheung, S.H. (2011). Structural Changes in the Lee-Carter Mortality Indexes: Detection and Implications. *North American Actuarial Journal*, **15**, 13-31.

Li, J.S.-H. and Hardy, M.R. (2011). Measuring Basis Risk in Longevity Hedges. *North American Actuarial Journal*, **15**, 177-200.

Li, J.S.-H., Hardy, M.R. and Tan, K.S. (2009). Uncertainty in Mortality Forecasting: An Extension to the Classical Lee-Carter Approach. *ASTIN Bulletin*, **39**, 137-164.

Li, J.S.-H. and Luo, A. (2012). Key q-Duration: A Framework for Hedging Longevity Risk. *ASTIN Bulletin*, **42**, 413-452.

Li, J.S.-H. and Ng, A.C.Y. (2011). Canonical Valuation of Mortality-Linked Securities. *Journal of Risk and Insurance*, **78**, 853-884.

Li, J.S.-H., Ng, A.C.Y. and Chan, W.S. (2011). On the Calibration of Mortality Forward Curves. *Journal of Futures Markets*, **31**, 941-970.

Li, J.S.-H., Zhou, R. and Hardy, M.R. (2015b). A Step-by-Step Guide to Building Multi-Population Stochastic Mortality Models. *Insurance: Mathematics and Economics*, in press.

Li, N., and Lee, R. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography*, **42**, 575-594.

Li, S.H. and Chan, W.S. (2005). Outlier Analysis and Mortality Forecasting: The United Kingdom and Scandinavian Countries. *Scandinavian Actuarial Journal*, **3**, 187-211.

Li, S.H. and Chan, W.S. (2007). The Lee-Carter Model for Forecasting Mortality Revisited. *North American Actuarial Journal*, **11**, 68-89.

Lin, Y. and Cox, S.H. (2008). Securitization of Catastrophe Mortality Risks. *Insurance: Mathematics and Economics*, **42**, 628-637.

Lin, Y., Liu, S. and Yu, J. (2013). Pricing Mortality Securities with Correlated Mortality Indexes. *Journal of Risk and Insurance*, **80**, 921-948.

Luciano, E., Regis, L., and Vigna, E. (2012). Delta-Gamma Hedging of Mortality and Interest Rate Risk. *Insurance: Mathematics and Economics*, **50**, 402-412.

Mavros, G., Cairns, A.J.G., Kleinow, T., and Streftaris, G. (2014). A Parsimonious Approach to Stochastic Mortality Modelling with Dependent Residuals. Working paper, Heriot-Watt University.

Michaelson, A. and Mulholland, J. (2014). Strategy for Increasing the Global Capacity for Longevity Risk Transfer: Developing Transactions That Attract Capital Markets Investors. *Journal of Alternative Investments*, **17**, 18-27.

Milevsky, M. (2005). Real Longevity Insurance with a Deductible: Introduction to Advanced-Life Delayed Annuities. *North American Actuarial Journal*, **9**, 109-122.

Milidonis, A., Lin, Y., and Cox, S.H. (2011). Mortality Regimes and Pricing. *North American Actuarial Journal*, **15**, 266-289.

Murphy, K.M. and Topel, R.H. (2002). Estimation and Inference in Two-Step Econometric Models. *Journal of Business and Economic Statistics*, **20**, 88-97.

Nyblom, J. and Mäkeläinen, T. (1983). Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model. *Journal of the American Statistical Association*, **384**, 856-864.

Ngai, A. and Sherris, M. (2011). Longevity Risk Management for Life and Variable Annuities: The Effectiveness of Static Hedging Using Longevity Bonds and Derivatives. *Insurance Mathematics and Economics*, **49**, 100-114.

Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. *Science*, **296**, 1029-1031.

O'Hare, C. and Li, Y. (2014). Identifying Structural Breaks in Stochastic Mortality Models. *Journal of Risk and Uncertainty in Engineering Part B*, Forthcoming.

Olivieri, A. and Pitacco, E. (2009). Stochastic Mortality: The Impact on Target Capital. *ASTIN Bulletin*, **39**, 541-563.

Pedroza, C. (2006). A Bayesian Forecasting Model: Predicting U.S. Male Mortality. *Biostatistics*, **7**, 530-550.

Plat, R. (2011). One-Year Value-at-Risk for Longevity and Mortality. *Insurance: Mathematics and Economics*, **49**, 462-470.

Potter, C.W. (2001). A History of Influenza. *Journal of Applied Microbiology*, **91**, 572-579.

Pyle, G.F. (1986). The Diffusion of Influenza: Patterns and Paradigms. New Jersey: Rowan and Littlefield.

Renshaw, A. and Haberman, S. (2003). Lee-Carter Mortality Forecasting: A Parallel Generalized Linear Modelling Approach for England and Wales Mortality Projections. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **52**, 119137.

Shumway, R.H and Stoffer, D.S. (2006). Time Series Analysis and Its Applications: With R Examples. Springer-Verlag: New York.

Society of Actuaries. (2014). Mortality Improvement Scale MP-2014 Report. Available at https://www.soa.org/Files/Research/Exp-Study/research-2014-mp-report.pdf.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. *Journal of the American Statistical Association*, **86**, 1024-1033.

Stracke, A. and Heinen, W. (2006). Influenza Pandemic: The Impact on an Insured Lives Life Insurance Portfolio. The Actuary June 2006.

Stutzer, M. (1996). A Simple Nonparametric Approach to Derivative Security Valuation. *Journal of Finance*, **51**, 1633-1652.

Sweeting, P.J. (2011). A Trend-Change Extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science*, **5**, 143-162.

Tan, C. I., Li, J., Li, J. S.-H. and Balasooriya, U. (2014). Parametric Mortality Indexes: From Index Construction to Hedging Strategies. *Insurance: Mathematics and Economics*, **59**, 285-299.

Tan, K.S., Blake, D. and MacMinn, R. (2015). Longevity Risk and Capital Markets: The 2013-14 Update. *Insurance: Mathematics and Economics*, **63**, 1-11.

Toole, J. (2007). Potential Impact of Pandemic Influenza on the U.S. Life Insurance Industry. Research Report, Society of Actuaries.
Available at http://www.soa.org/Files/Research/Projects/resrch-li-rep-pan-life.pdf.

Van Berkum, F., Antonio, K., and Vellekoop, M. (2014). The Impact of Multiple Structural Changes on Mortality Predictions. *Scandinavian Actuarial Journal*, doi: 10.1080/03461238.2014.987807.

Vaupel, J.W. (1997). The Remarkable Improvements in Survival at Older Ages. *Philosophical transactions of the Royal Society of London, B*, **352**, 1799–1804.

Wang, C.-W., Yang, S.S. and Huang, H.-C. (2015) Modeling Multi-Country Mortality Dependence and Its Application in Pricing Survivor Index Swaps–A Dynamic Copula Approach. *Insurance: Mathematics and Economics*, **63**, 30-39.

Webster, R.G., Shortridge, K.F. and Kawaoka, Y. (1997). Influenza: Interspecies Transmission and Emergence of New Pandemics. *FEMS Immunology and Medical Microbiology*, **18**, 275-279.

Wills, S. and Sherris, M. (2010). Securitization, Structuring and Pricing of Longevity Risk. *Insurance: Mathematics and Economics*, **46**, 173-185.

Wilmoth, J.R., Andreev, E., Jdanov, D.A. and Glei, D.A. (2005). Methods Protocol for the Human Mortality Database. www.mortality.org.

Yang, B., Li, J. and Balasooriya, U. (2015). Using Bootstrapping to Incorporate Model Error for Risk-Neutral Pricing of Longevity Risk. *Insurance: Mathematics and Economics*, **62**, 16-27.

Yang, S.S. and Wang, C.W. (2013). Pricing and Securitization of Multi-Country Longevity Risk with Mortality Dependence. *Insurance: Mathematics and Economics*, **52**, 157-169.

Zhou, K.Q. and Li, J.S.-H. (2014). Dynamic Longevity Hedging in the Presence of Population Basis Risk: A Feasibility Analysis from Technical and Economic Perspectives. Paper presented at the 10th International Longevity Risk and Capital Markets Solutions Conference, Santiago, Chile.

Zhou, R. and Li, J.S.H. (2013). A Cautionary Note on Pricing Longevity Index Swaps. *Scandinavian Actuarial Journal*, **2013**, 1-23.

Zhou, R., Li, J.S.-H. and Tan, K.S. (2011). Economic Pricing of Mortality-Linked Securities in the Presence of Population Basis Risk. *Geneva Paper of Risk and Insurance: Issues and Practice*, **36**, 544-566.

Zhou, R., Li, J.S.-H. and Tan, K.S. (2013a). Pricing Mortality Risk: A Two-Population Model with Transitory Jump Effects. *Journal of Risk and Insurance*, **80**, 733-774.

Zhou, R., Li, J.S.-H. and Tan, K.S. (2013b). Economic Pricing of Mortality-Linked Securities: A Tâtonnement Approach. *Journal of Risk and Insurance*, doi: 10.1111/j.1539-6975.2013.12008.x.

Zhou, R., Wang, Y., Kaufhold, K., Li, J.S.-H. and Tan, K.S. (2014). Modeling Period Effects in Multi-Population Mortality Models: Applications to Solvency II. *North American Actuarial Journal*, **18**, 150-167.

Appendix A

A.1 Derivation of Property 1, 2 and 3

The assumptions required in the derivation of Property 1, 2 and 3 are summarized as follow:

- $\xi_t \sim \mathrm{N}(0, \sigma^2), \vec{J}_t \sim \mathrm{MVN}(\vec{\mu}^{(J)}, \Sigma_J), \vec{\varepsilon}_t \sim \mathrm{MVN}(0, \mathrm{I} \cdot \sigma_r^2)$, and
- $\xi_t, N_t, \vec{J}_t$ and $\vec{\varepsilon}_t$ are mutually independent, and
- ξ_t, N_t and $\vec{J}_t$ are serially independent.

Derivation of Property 1:

According to equation (2.6), we have

$$\vec{Z}_t = \vec{b}(\mu + \xi_t) + N_t\vec{J}_t - N_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t,$$

$$\vec{Z}_{t+1} = \vec{b}(\mu + \xi_{t+1}) + N_{t+1}\vec{J}_{t+1} - N_t\vec{J}_t + \vec{\varepsilon}_{t+1}.$$

Therefore, $\vec{Z}_t$ given $N_t = n_t, \vec{J}_t = \vec{j}_t$ would be $\vec{b}(\mu+\xi_t)+n_t\vec{j}_t-N_{t-1}\vec{J}_{t-1}+\vec{\varepsilon}_t$. Similarly, $\vec{Z}_{t+1}$ given $N_t = n_t, \vec{J}_t = \vec{j}_t$ would be $\vec{b}(\mu+\xi_{t+1})+N_{t+1}\vec{J}_{t+1}-n_t\vec{j}_t+\vec{\varepsilon}_{t+1}$. Since that we have the assumption of $\xi_t, N_t, \vec{J}_t$ and $\vec{\varepsilon}_t$ being mutually and serially independent, we immediately get

$$\vec{Z}_t|N_t = n_t, \vec{J}_t = \vec{j}_t \quad \perp \quad \vec{Z}_{t+1}|N_t = n_t, \vec{J}_t = \vec{j}_t.$$

Derivation of Property 2:

As shown in equation (2.6), we have

$$\vec{Z}_t = \vec{b}(\mu + \xi_t) + N_t\vec{J}_t - N_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t.$$

Therefore, $\vec{Z}_t$ given $N_{t-1} = n_{t-1}, N_t = n_t$ would be $\vec{b}(\mu + \xi_t) + n_t\vec{J}_t - n_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t$.

The moment generating function of $\vec{Z}_t$ given $N_{t-1} = n_{t-1}, N_t = n_t$ can then be computed as

$$\begin{aligned} M_{\vec{Z}_t|N_{t-1}=n_{t-1},N_t=n_t}(\vec{x}) &= \mathrm{E}(\exp(\vec{x}'\vec{Z}_t)|N_t = n_t, N_{t-1} = n_{t-1}) \\ &= \mathrm{E}(\exp(\vec{x}'(\vec{b}(\mu + \xi_t) + n_t\vec{J}_t - n_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t))) \\ &= \mathrm{E}(\exp(\vec{x}'\vec{b}(\mu + \xi_t))) \cdot \mathrm{E}(\exp(\vec{x}'n_t\vec{J}_t)) \cdot \mathrm{E}(\exp(-\vec{x}'n_{t-1}\vec{J}_{t-1})) \cdot \mathrm{E}(\exp(\vec{x}'\varepsilon_t)). \end{aligned}$$

Since that we have

- $\xi_t \sim \mathrm{N}(0, \sigma^2), \vec{J}_t \sim \mathrm{MVN}(\vec{\mu}^{(J)}, \Sigma_J), \vec{\varepsilon}_t \sim \mathrm{MVN}(0, \mathrm{I} \cdot \sigma_r^2)$, and
- $\xi_t, \vec{J}_t$ and $\vec{\varepsilon}_t$ are mutually independent, and
- $\vec{J}_t$ are serially independent,

the moment generating function of $\vec{Z}_t$ given $N_{t-1} = n_{t-1}, N_t = n_t$ can then be computed as

$$\begin{aligned} M_{\vec{Z}_t|N_{t-1}=n_{t-1},N_t=n_t}(\vec{x}) &= \exp(\vec{x}'\vec{b}\mu + \tfrac{1}{2}\vec{x}'\vec{b}\sigma^2\vec{b}'\vec{x})) \cdot \exp(\vec{x}'n_t\vec{\mu}^{(J)} + \tfrac{1}{2}\vec{x}'n_t^2\Sigma_J\vec{x}) \\ &\quad \cdot \exp(-\vec{x}'n_{t-1}\vec{\mu}^{(J)} + \tfrac{1}{2}\vec{x}'n_{t-1}^2\Sigma_J\vec{x}) \cdot \exp(\tfrac{1}{2}\vec{x}'\mathrm{I}\sigma_r^2\vec{x}) \\ &= \exp(\vec{x}'(\vec{b}\mu + n_t\vec{\mu}^{(J)} - n_{t-1}\vec{\mu}^{(J)}) + \tfrac{1}{2}\vec{x}'(\vec{b}\vec{b}'\sigma^2 + n_t^2\Sigma_J + n_{t-1}^2\Sigma_J + \mathrm{I}\sigma_r^2)\vec{x}) \\ &= \exp(\vec{x}'\vec{\mu}_{mar} + \tfrac{1}{2}\vec{x}'\Sigma_{mar}\vec{x}), \end{aligned}$$

where

$$\vec{\mu}_{mar} = \vec{b}\mu + (n_t - n_{t-1})\vec{\mu}^{(J)}$$

and

$$\Sigma_{mar} = \vec{b}\vec{b}'\sigma^2 + \mathrm{I} \cdot \sigma_r^2 + (n_t^2 + n_{t-1}^2)\Sigma_J.$$

Therefore,

$$\vec{Z}_t | N_{t-1} = n_{t-1}, N_t = n_t \sim MVN(\vec{\mu}_{mar}, \Sigma_{mar}),$$

where

$$\vec{\mu}_{mar} = \vec{b}\mu + (n_t - n_{t-1})\vec{\mu}^{(J)}$$

and

$$\Sigma_{mar} = \vec{b}\vec{b}'\sigma^2 + \mathrm{I} \cdot \sigma_r^2 + (n_t^2 + n_{t-1}^2)\Sigma_J.$$

Derivation of Property 3:

According to equation (2.6), we have

$$\vec{Z}_t = \vec{b}(\mu + \xi_t) + N_t\vec{J}_t - N_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t,$$

$$\vec{Z}_{t+1} = \vec{b}(\mu + \xi_{t+1}) + N_{t+1}\vec{J}_{t+1} - N_t\vec{J}_t + \vec{\varepsilon}_{t+1}.$$

Define

$$\vec{Z}^*_{t,t+1} = \begin{pmatrix} \vec{Z}_t \\ \vec{Z}_{t+1} \end{pmatrix}.$$

Then $\vec{Z}^*_{t,t+1}$ given $N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1}$ can be expressed as

$$\begin{pmatrix} \vec{b}(\mu + \xi_t) + n_t\vec{J}_t - n_{t-1}\vec{J}_{t-1} + \vec{\varepsilon}_t \\ \vec{b}(\mu + \xi_{t+1}) + n_{t+1}\vec{J}_{t+1} - n_t\vec{J}_t + \vec{\varepsilon}_{t+1} \end{pmatrix}.$$

To enhance the derivation of Property 3, we impose a working independence assumption that the error term $\vec{\varepsilon}_t$ is serially independent. Based on this working independence assumption, the covariance of $\vec{Z}_t$ and $\vec{Z}_{t+1}$ given $N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1}$ would equal to $-n_t^2\Sigma_J$. The resulting likelihood function would then become a composite likelihood (e.g., see Diao and Cook, 2014), which is a partial specification of the full likelihood.

Using Property 2, we have

$$\vec{Z}^*_{t,t+1} | N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1} \sim MVN(\vec{\mu}_{joint}, \Sigma_{joint}),$$

where

$$\vec{\mu}_{joint} = \begin{pmatrix} \vec{b}\mu + (n_t - n_{t-1})\vec{\mu}^{(J)} \\ \vec{b}\mu + (n_{t+1} - n_t)\vec{\mu}^{(J)} \end{pmatrix}$$

and

$$\Sigma_{joint} = \left(\begin{array}{c|c} \sigma^2 \vec{b}\vec{b}' + \mathrm{I} \cdot \sigma_r^2 + (n_t^2 + n_{t-1}^2)\Sigma_J & -n_t^2 \Sigma_J \\ \hline -n_t^2 \Sigma_J & \sigma^2 \vec{b}\vec{b}' + \mathrm{I} \cdot \sigma_r^2 + (n_{t+1}^2 + n_t^2)\Sigma_J \end{array} \right).$$

A.2 Estimation Algorithms for Model J0, J1 and J2

In this appendix, we present the algorithms for estimating the three model variants under consideration. We first provide in Appendix A.2.1 the information that applies to all three model variants. Then in Appendices A.2.2 to A.2.4 we give the details that are specific to each model variant.

A.2.1 General Information

Recall that the log-likelihood function is given by

$$l(\vec{\theta}) = \ln f(\vec{z}_{t_2}, \ldots, \vec{z}_{t_T}; \vec{\theta}) = \sum_{t=t_2}^{t_{T-1}} \ln f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta}) - \sum_{t=t_3}^{t_{T-1}} \ln f(\vec{z}_t; \vec{\theta}).$$

Parameter estimates are obtained by maximizing $l(\vec{\theta})$ with an iterative Newton-Raphson procedure, in which parameters are updated one at a time. The updating of a parameter θ in the parameter vector $\vec{\theta}$ proceeds according to

$$\theta^{(m+1)} = \theta^{(m)} - \frac{\frac{\partial}{\partial\theta} l(\vec{\theta})|_{\theta=\theta^{(m)}}}{\frac{\partial^2}{\partial\theta^2} l(\vec{\theta})|_{\theta=\theta^{(m)}}}, \tag{1}$$

where $\theta^{(m)}$ represents the estimate of θ in the mth iteration. The identifiability constraints are applied at the end of each iteration.

Equation (1) depends on the first and second partial derivatives of $l(\vec{\theta})$ with respect to

θ, which are given by

$$\frac{\partial}{\partial\theta}l(\vec{\theta}) = \sum_{t=t_2}^{t_{T-1}} \frac{\frac{\partial}{\partial\theta}f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})}{f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})} - \sum_{t=t_3}^{t_{T-1}} \frac{\frac{\partial}{\partial\theta}f(\vec{z}_t; \vec{\theta})}{f(\vec{z}_t; \vec{\theta})}$$

and

$$\frac{\partial^2}{\partial\theta^2}l(\theta) = \sum_{t=t_2}^{t_{T-1}} \Big[\frac{\frac{\partial^2}{\partial\theta^2}f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})}{f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})} - \Big(\frac{\frac{\partial}{\partial\theta}f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})}{f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})}\Big)^2\Big] - \sum_{t=t_3}^{t_{T-1}} \Big[\frac{\frac{\partial^2}{\partial\theta^2}f(\vec{z}_t; \vec{\theta})}{f(\vec{z}_t; \vec{\theta})} - \Big(\frac{\frac{\partial}{\partial\theta}f(\vec{z}_t; \vec{\theta})}{f(\vec{z}_t; \vec{\theta})}\Big)^2\Big],$$

respectively. The expressions above involve the first and second partial derivatives of $f(\vec{z}_t, \vec{z}_{t+1}; \vec{\theta})$ and $f(\vec{z}_t; \vec{\theta})$ with respect to θ. These partial derivatives are derived as follows.

First, we consider the following conditional density functions:

$$f_t^{n_{t-1},n_t} := f(\vec{z}_t | N_{t-1} = n_{t-1}, N_t = n_t),$$

$$f_{t,t+1}^{n_{t-1},n_t,n_{t+1}} := f(\vec{z}_t, \vec{z}_{t+1} | N_{t-1} = n_{t-1}, N_t = n_t, N_{t+1} = n_{t+1}).$$

The first and second partial derivatives of $f_t^{n_{t-1},n_t}$ with respect to θ can be expressed as

$$\begin{aligned}\tfrac{\partial}{\partial\theta}f_t^{n_{t-1},n_t} &= f_t^{n_{t-1},n_t} \cdot [-\tfrac{1}{2} \cdot \mathrm{Tr}(\Sigma_{mar}^{-1} \cdot \tfrac{\partial}{\partial\theta}\Sigma_{mar}) - \tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar})^T\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar}) \\ &\quad -\tfrac{1}{2} \cdot (\vec{z}_t - \vec{\mu}_{mar})^T \tfrac{\partial}{\partial\theta}\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar})]\end{aligned}$$

and

$$\begin{aligned}\tfrac{\partial^2}{\partial\theta^2}f_t^{n_{t-1},n_t} &= f_t^{n_{t-1},n_t} \cdot \{[-\tfrac{1}{2}\mathrm{Tr}(\Sigma_{mar}^{-1} \cdot \tfrac{\partial}{\partial\theta}\Sigma_{mar}) - \tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar})^T\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar}) \\ &\quad -\tfrac{1}{2}(\vec{z}_t - \vec{\mu}_{mar})^T \tfrac{\partial}{\partial\theta}\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar})]^2 - \tfrac{1}{2}\mathrm{Tr}(\tfrac{\partial}{\partial\theta}\Sigma_{mar}^{-1}\tfrac{\partial}{\partial\theta}\Sigma_{mar} + \Sigma_{mar}^{-1}\tfrac{\partial^2}{\partial\theta^2}\Sigma_{mar}) \\ &\quad -\tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar})^T \tfrac{\partial}{\partial\theta}\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar}) - \tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar})^T\Sigma_{mar}^{-1}\tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar}) \\ &\quad -\tfrac{\partial}{\partial\theta}(\vec{z}_t - \vec{\mu}_{mar})^T \tfrac{\partial}{\partial\theta}\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar}) - \tfrac{1}{2}(\vec{z}_t - \vec{\mu}_{mar})^T \tfrac{\partial^2}{\partial\theta^2}\Sigma_{mar}^{-1}(\vec{z}_t - \vec{\mu}_{mar})\},\end{aligned}$$

respectively, where $\mathrm{Tr}(M)$ denotes the trace of a matrix M. The first and second partial derivatives of $f_{t,t+1}^{n_{t-1},n_t,n_{t+1}}$ with respect to θ can be obtained by replacing $\vec{z}_t$, $\vec{\mu}_{mar}$ and Σ_{mar} in the expressions above by $\vec{z}_{t,t+1}^{\star}$, $\vec{\mu}_{joint}$ and Σ_{joint}, respectively.

Then, by using the facts that

$$\begin{aligned} f(\vec{z}_t;\vec{\theta}) &= p^2 f_t^{11} + p(1-p) f_t^{10} + p(1-p) f_t^{01} + (1-p)^2 f_t^{00} \\ &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} p^{n_{t-1}+n_t} \cdot (1-p)^{2-(n_{t-1}+n_t)} \cdot f_t^{n_{t-1},n_t} \end{aligned}$$

and that

$$\begin{aligned} f(\vec{z}_t, \vec{z}_{t+1};\vec{\theta}) &= p \cdot f(\vec{z}_t, \vec{z}_{t+1}|N_t=1) + (1-p) \cdot f(\vec{z}_t, \vec{z}_{t+1}|N_t=0) \\ &= p^3 \cdot f_{t,t+1}^{111} + p^2(1-p) \cdot [f_{t,t+1}^{110} + f_{t,t+1}^{011} + f_{t,t+1}^{101}] \\ &\quad + p(1-p)^2 \cdot [f_{t,t+1}^{010} + f_{t,t+1}^{100} + f_{t,t+1}^{001}] + (1-p)^3 \cdot f_{t,t+1}^{000} \\ &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} \sum_{n_{t+1}=0}^{1} p^{n_{t-1}+n_t+n_{t+1}} \cdot (1-p)^{3-(n_{t-1}+n_t+n_{t+1})} \cdot f_{t,t+1}^{n_{t-1},n_t,n_{t+1}}, \end{aligned}$$

we can readily compute the first and second partial derivatives of $f(\vec{z}_t, \vec{z}_{t+1};\vec{\theta})$ and $f(\vec{z}_t;\vec{\theta})$ with respect to θ as described below.

Case I: $\theta \neq p$

$$\begin{aligned} \tfrac{\partial}{\partial\theta} f(\vec{z}_t;\vec{\theta}) &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} p^{n_{t-1}+n_t} \cdot (1-p)^{2-(n_{t-1}+n_t)} \cdot \frac{\partial}{\partial\theta} f_t^{n_{t-1},n_t} \\ \tfrac{\partial^2}{\partial\theta^2} f(\vec{z}_t;\vec{\theta}) &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} p^{n_{t-1}+n_t} \cdot (1-p)^{2-(n_{t-1}+n_t)} \cdot \frac{\partial^2}{\partial\theta^2} f_t^{n_{t-1},n_t} \end{aligned}$$

$$\begin{aligned} \tfrac{\partial}{\partial\theta} f(\vec{z}_t, \vec{z}_{t+1};\vec{\theta}) &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} \sum_{n_{t+1}=0}^{1} p^{n_{t-1}+n_t+n_{t+1}} \cdot (1-p)^{3-(n_{t-1}+n_t+n_{t+1})} \cdot \frac{\partial}{\partial\theta} f_{t,t+1}^{n_{t-1},n_t,n_{t+1}} \\ \tfrac{\partial^2}{\partial\theta^2} f(\vec{z}_t, \vec{z}_{t+1};\vec{\theta}) &= \sum_{n_{t-1}=0}^{1} \sum_{n_t=0}^{1} \sum_{n_{t+1}=0}^{1} p^{n_{t-1}+n_t+n_{t+1}} \cdot (1-p)^{3-(n_{t-1}+n_t+n_{t+1})} \cdot \frac{\partial^2}{\partial\theta^2} f_{t,t+1}^{n_{t-1},n_t,n_{t+1}} \end{aligned}$$

Case II: $\theta = p$

$$\begin{aligned}
\tfrac{\partial}{\partial\theta} f(\vec{z}_t;\vec{\theta}) &= 2p\cdot f_t^{11} + (1-2p)\cdot f_t^{10} + (1-2p)\cdot f_t^{01} - 2(1-p) f_t^{00} \\
\tfrac{\partial^2}{\partial\theta^2} f(\vec{z}_t;\vec{\theta}) &= 2f_t^{11} - 2f_t^{10} - 2f_t^{01} + 2f_t^{00}
\end{aligned}$$

$$\begin{aligned}
\tfrac{\partial}{\partial\theta} f(\vec{z}_t,\vec{z}_{t+1};\vec{\theta}) &= 3p^2\cdot f_{t,t+1}^{111} + (2p-3p^2)\cdot[f_{t,t+1}^{110} + f_{t,t+1}^{011} + f_{t,t+1}^{101}] \\
&\quad +(1-4p+3p^2)\cdot[f_{t,t+1}^{010} + f_{t,t+1}^{100} + f_{t,t+1}^{001}] - 3(1-p)^2\cdot f_{t,t+1}^{000} \\
\tfrac{\partial^2}{\partial\theta^2} f(\vec{z}_t,\vec{z}_{t+1};\vec{\theta}) &= 6p\cdot f_{t,t+1}^{111} + (2-6p)\cdot[f_{t,t+1}^{110} + f_{t,t+1}^{011} + f_{t,t+1}^{101}] \\
&\quad +(-4+6p)\cdot[f_{t,t+1}^{010} + f_{t,t+1}^{100} + f_{t,t+1}^{001}] + 6(1-p)\cdot f_{t,t+1}^{000}
\end{aligned}$$

A.2.2 Model J0

The parameters needed to be estimated include:

$$\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_X \end{pmatrix}, \quad \mu, \quad \ln\sigma, \quad \ln\sigma_r, \quad \mu_J, \quad \ln\sigma_J, \quad \text{logit}(p).$$

We estimate σ, σ_r and σ_J in natural log scale and p in logit scale to avoid any potential boundary problems. The initial value for each parameter is chosen as follows:

$\vec{b}$: the maximum likelihood estimates of $\vec{b}$ in the original Lee-Carter model;

μ: the mean of the first difference of the sequence of the maximum likelihood estimates of k_t in the original Lee-Carter model;

$\ln\sigma$: the sample standard deviation (in log scale) of the sequence of the maximum likelihood estimates of k_t in the original Lee-Carter model;

$\ln\sigma_r$: the maximum likelihood estimate of σ_r (in log scale) in the original Lee-Carter model, fitted using the Route II approach;

μ_J: same as the initial value for μ;

$\ln \sigma_J$: same as the initial value for $\ln \sigma$;

$\text{logit}(p)$: the frequency of the detected additive outliers (in logit scale).

Given the initial values, parameters are updated iteratively by equation (1), until the change in the value of the log-likelihood function is smaller than a tolerance level, which we set to 10^{-8}.

A.2.3 Model J1

The parameters needed to be estimated include:

$$\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_X \end{pmatrix}, \quad \mu, \quad \ln \sigma, \quad \ln \sigma_r, \quad \mu_J, \quad \ln \sigma_J, \quad \vec{b}^{(J)} = \begin{pmatrix} b_1^{(J)} \\ b_2^{(J)} \\ \vdots \\ b_X^{(J)} \end{pmatrix}, \quad \text{logit}(p).$$

The initial value for each parameter is chosen as follows:

$\vec{b}$: the maximum likelihood estimate of $\vec{b}$ in Model J0;

μ: the maximum likelihood estimate of μ in Model J0;

$\ln \sigma$: the maximum likelihood estimate of $\ln \sigma$ in Model J0;

$\ln \sigma_r$: the maximum likelihood estimate of $\ln \sigma_r$ in Model J0;

μ_J: the maximum likelihood estimate of μ_J in Model J0;

$\ln \sigma_J$: the maximum likelihood estimate of $\ln \sigma_J$ in Model J0;

$\vec{b}^{(J)}$: the maximum likelihood estimate of $\vec{b}$ in Model J0;

$\text{logit}(p)$: the maximum likelihood estimate of $\text{logit}(p)$ in Model J0;.

As before, parameters are updated iteratively using equation (1), until the change in the value of the log-likelihood function is smaller than the tolerance level of 10^{-8}.

A.2.4 Model J2

The parameters needed to be estimated include:

$$\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_X \end{pmatrix}, \quad \mu, \quad \ln\sigma, \quad \ln\sigma_r, \quad \vec{b}^{(J)} = \begin{pmatrix} b_1^{(J)} \\ b_2^{(J)} \\ \vdots \\ b_X^{(J)} \end{pmatrix}, \quad \ln\sigma_J, \quad \rho, \quad \text{logit}(p).$$

Compared to Model J1, this model variant contains only one extra parameter, ρ, which determines the correlation between the jump effects on two different age groups. From Figure 2.1, we observe that the shapes of the age patterns of the historical mortality jumps are quite stable. It is therefore reasonable to expect that ρ is fairly close to (and smaller than) one. This expectation motivates us to use the profile likelihood technique to expedite convergence.[1] The overall estimation procedure is summarized as follows:

1. Partition the parameter vector $\vec{\theta}$ into two groups, namely the correlation parameter ρ and other parameters $\vec{\theta}^*$. Let $l(\vec{\theta}^*|\rho)$ be the model's log-likelihood conditioned on ρ.

2. Starting with $\rho = 0.99999$, estimate $\vec{\theta}^*$ by maximizing $l(\vec{\theta}^*|\rho)$ with the iterative Newton-Raphson procedure. The required initial values are taken as the corresponding maximum likelihood estimates in Model J1. The maximized value of $l(\vec{\theta}^*|\rho)$ is recorded.

3. Reduce ρ by 0.00001 and repeat Step 2. Stop the algorithm when the maximized value of $l(\vec{\theta}^*|\rho)$ no longer increases.

A.3 The Parametric Bootstrap for Model J0, J1 and J2

To measure parameter uncertainty, we use a parametric bootstrapping procedure that is adapted from the work of Brouhns et al. (2005). The procedure used is summarized as follows:

[1] We refer interested readers to Davison (2003) for further information about the profile likelihood technique.

1. For each t in the data sample period, simulate a realization of each of the following:
   - the jump indicator variable N_t from $B(1, \hat{p})$;
   - the innovation term ξ_t in the random walk with drift from $N(0, \hat{\sigma}^2)$;
   - the vector of jump effects $\vec{J}_t$ from $MVN(\hat{\vec{\mu}}_J, \hat{\Sigma}_J)$, provided that the simulated value of N_t is 1.
2. For each x in the sample age range and each t in the data sample period, simulate a realization of the residual $\varepsilon_{x,t}$ from $N(0, \hat{\sigma}_r^2)$.
3. Calculate realizations of $Z_{x,t}$ for all x in the sample age range and t in the data sample period using equation (2.5) and the results from the previous two steps.
4. On the basis of the realizations of $Z_{x,t}$ from the previous step, re-estimate all model parameters by the method of maximum likelihood, as described in Appendix A.2.
5. Repeat the previous four steps M times.

In the above, the "hat" sign over a parameter denotes a maximum likelihood estimate from the original data sample.

The bootstrapping procedure provides us with M collections of parameter estimates, from which the standard error of each parameter estimate can be calculated. We use $M = 5,000$ in our calculations. Note that the standard errors of $\vec{b}^{(J)}$, $\ln \sigma_J$ and ρ are computed from a subset of the 5,000 realizations in which at least one simulated value of N_t is non-zero. In addition, the standard errors of $\ln \sigma$, $\ln \sigma_r$, $\ln \sigma_J$ and $\text{logit}(p)$ are computed from realizations in their original scales (i.e., after an anti-log or anti-logit transformation, where applicable).

Appendix B

B.1 Estimation Procedure for the LLCBD model

In this appendix we detail the estimation procedure, which is adapted from the work of Holmes (2013). As before, we use $[t_a, t_b]$ and $[x_a, x_b]$ to denote the sample period and sample age range to which the model is fitted, respectively. The vector of observations at time t is $\vec{y}_t = (y_{x_a,t}, \ldots, y_{x_b,t})'$.

Following Holmes (2013), we use an Expectation-Maximization (EM) algorithm to obtain maximum likelihood parameter estimates. Under our model assumptions,

$$\vec{\eta}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, Q) \quad \text{and} \quad \vec{\epsilon}_t \overset{\text{i.i.d.}}{\sim} \text{MVN}(0, \text{I}_{x_b - x_a + 1}\sigma_\epsilon^2).$$

It immediately follows that the log-likelihood function is given by

$$\begin{aligned}\ln(\mathcal{L}) = &-\tfrac{1}{2\sigma_\epsilon^2} \sum_{t=t_a}^{t_b} (\vec{y}_t - B\vec{\alpha}_t)' (\vec{y}_t - B\vec{\alpha}_t) - \tfrac{(t_b - t_a + 1)(x_b - x_a + 1)}{2} \ln(\sigma_\epsilon^2) \\ &-\tfrac{1}{2} \sum_{t=t_a+1}^{t_b} (\vec{\alpha}_t - A\vec{\alpha}_{t-1})' Q^{-1} (\vec{\alpha}_t - A\vec{\alpha}_{t-1}) - \tfrac{1}{2} \sum_{t=t_a+1}^{t_b} \ln|Q| + c_l,\end{aligned}$$

where c_l is a constant that is free of the hidden states and parameters. When fitting the special case with constant drifts, $\vec{\alpha}_t$ and Q should be replaced by $\vec{\alpha}_t^*$ and Q^*, respectively.

The EM algorithm iterates over two steps until convergence. In the first step, which is known as the Expectation step, the expectation of the log-likelihood is computed:

$$\begin{aligned}\Psi &= \text{E}\left[\ln(\mathcal{L})\right] \\ &= -\tfrac{1}{2\sigma_\epsilon^2} \sum_{t=t_a}^{t_b} \Big(\text{E}[\vec{y}_t'\vec{y}_t] - 2\text{E}[\vec{y}_t' B\vec{\alpha}_t] + \text{E}[\vec{\alpha}_t' B' B\vec{\alpha}_t]\Big) - \tfrac{(t_b - t_a + 1)(x_b - x_a + 1)}{2} \ln(\sigma_\epsilon^2) \\ &\quad -\tfrac{1}{2} \sum_{t=t_a+1}^{t_b} \Big(\text{E}[\vec{\alpha}_t' Q^{-1} \vec{\alpha}_t] - 2\text{E}[\vec{\alpha}_{t-1}' A' Q^{-1} \vec{\alpha}_t] + \text{E}[\vec{\alpha}_{t-1}' A' Q^{-1} A\vec{\alpha}_{t-1}]\Big) - \tfrac{t_b - t_a}{2} \ln|Q| + c_l.\end{aligned}$$

In the second step, which is known as the Maximization step, parameter estimates are obtained by maximizing the expected log-likelihood Ψ. We now derive the update equations for parameters σ_ϵ^2 and Q in the Maximization step.

Update Equation for σ_ϵ^2:
Differentiating Ψ with respect to σ_ϵ^2, we have

$$\frac{\partial \Psi}{\partial \sigma_\epsilon^2} = -\frac{1}{2}\sum_{t=t_a}^{t_b} \frac{-\mathrm{E}\left[\vec{y}_t'\vec{y}_t\right] + 2\mathrm{E}\left[\vec{y}_t' B\vec{\alpha}_t\right] - \mathrm{E}\left[\vec{\alpha}_t' B' B\vec{\alpha}_t\right]}{\sigma_\epsilon^4} - \frac{(t_b - t_a + 1)(x_b - x_a + 1)}{2\sigma_\epsilon^2}.$$

Setting the partial derivative to 0, we can obtain the update equation for R as follows:

$$\begin{aligned}
\sigma_\epsilon^2 &= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \left(\mathrm{E}\left[\vec{y}_t'\vec{y}_t\right] - 2\mathrm{E}\left[\vec{y}_t' B\vec{\alpha}_t\right] + \mathrm{E}\left[\vec{\alpha}_t' B' B\vec{\alpha}_t\right]\right) \\
&= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \mathrm{E}\left[(\vec{y}_t - B\vec{\alpha}_t)'(\vec{y}_t - B\vec{\alpha}_t)\right] \\
&= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \mathrm{E}[\mathrm{vec}((\vec{y}_t - B\vec{\alpha}_t)'\mathrm{I}_{x_b - x_a + 1}(\vec{y}_t - B\vec{\alpha}_t))] \\
&= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \mathrm{E}\left[(\vec{y}_t - B\vec{\alpha}_t)' \otimes (\vec{y}_t - B\vec{\alpha}_t)'\right] \mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right) \\
&= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \mathrm{E}\left[\mathrm{vec}((\vec{y}_t - B\vec{\alpha}_t)(\vec{y}_t - B\vec{\alpha}_t)')'\right] \mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right) \\
&= \tfrac{1}{u}\sum_{t=t_a}^{t_b} \left(\mathrm{vec}(\mathrm{E}[\vec{y}_t\vec{y}_y'])' - 2\mathrm{vec}(\mathrm{E}[\vec{y}_t\vec{\alpha}_t']B')' + \mathrm{vec}(B\mathrm{E}[\vec{\alpha}_t\vec{\alpha}_t']B')'\right) \mathrm{vec}\left(\mathrm{I}_{x_b - x_a + 1}\right),
\end{aligned}$$

where $\mathrm{vec}(X)$ represents the vectorization of X, $\otimes$ is the kronecker product operator and $u = (t_b - t_a + 1)(x_b - x_a + 1)$.

Update Equation for Q:
The partial derivative of Ψ with respect to Q can be calculated as

$$\frac{\partial \Psi}{\partial Q} = \frac{1}{2}\sum_{t=t_a+1}^{t_b} Q^{-1}\Big(\mathrm{E}[\vec{\alpha}_t\vec{\alpha}_t'] - \mathrm{E}[\vec{\alpha}_t\vec{\alpha}_{t-1}']A' - A\mathrm{E}[\vec{\alpha}_{t-1}\vec{\alpha}_t'] + A\mathrm{E}[\vec{\alpha}_{t-1}\vec{\alpha}_{t-1}']A'\Big)Q^{-1} - \tfrac{t_b - t_a}{2}Q^{-1}.$$

Setting the partial derivative to zero, we obtain the following update equation for Q:

$$Q = \frac{1}{t_b - t_a} \sum_{t=t_a+1}^{t_b} \Big(\mathrm{E}[\vec{\alpha}_t \vec{\alpha}_t'] - \mathrm{E}[\vec{\alpha}_t \vec{\alpha}_{t-1}'] A' - A\mathrm{E}[\vec{\alpha}_{t-1} \vec{\alpha}_t'] + A\mathrm{E}[\vec{\alpha}_{t-1} \vec{\alpha}_{t-1}'] A' \Big).$$

Calculating the Expectations in the Update Equations:
We now explain how the expectations in the update equations for σ_ϵ^2 and Q can be evaluated. Let $\mathrm{E}_t(\cdot)$, $\mathrm{Var}_t(\cdot)$, $\mathrm{Cov}_t(\cdot, \cdot)$ respectively be the expectation, variance and covariance conditioned on the information up to and including time t. The expectations in the update equations are calculated on the basis of all data, so that, for example, $\mathrm{E}[\vec{\alpha}_t \vec{\alpha}_t']$ is computed as $\mathrm{E}_{t_b}[\vec{\alpha}_t \vec{\alpha}_t']$.

To calculate $\mathrm{E}_{t_b}[\vec{\alpha}_t \vec{\alpha}_{t-1}']$, $\mathrm{E}_{t_b}[\vec{\alpha}_t \vec{\alpha}_t']$, $\mathrm{E}_{t_b}[\vec{\alpha}_{t-1} \vec{\alpha}_t']$ and $\mathrm{E}_{t_b}[\vec{\alpha}_{t-1} \vec{\alpha}_{t-1}']$, we use the Kalman smoother algorithm. The first step in the algorithm is to calculate the expectations of $\vec{\alpha}_t \vec{\alpha}_t'$ and $\vec{\alpha}_t$, conditioned on the information up to an including time t, using the Kalman filter:

$$\begin{cases} \mathrm{E}_t[\vec{\alpha}_t] = \mathrm{E}_{t-1}[\vec{\alpha}_t] + \mathcal{K}_t(y_t - B\mathrm{E}_{t-1}[\vec{\alpha}_t]) \\ \mathrm{Var}_t(\vec{\alpha}_t) = (\mathrm{I}_m - \mathcal{K}_t B)\mathrm{Var}_{t-1}(\vec{\alpha}_t) \\ \mathrm{E}_t[\vec{\alpha}_t \vec{\alpha}_t'] = \mathrm{Var}_t(\vec{\alpha}_t) + \mathrm{E}_t[\vec{\alpha}_t](\mathrm{E}_t[\vec{\alpha}_t])' \end{cases},$$

where $\mathrm{E}_{t-1}[\vec{\alpha}_t] = A\mathrm{E}_{t-1}[\vec{\alpha}_{t-1}]$, $\mathrm{Var}_{t-1}(\vec{\alpha}_t) = A\mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})A' + Q$ and

$$\mathcal{K}_t = \mathrm{Var}_{t-1}(\vec{\alpha}_t)B' \left(B\mathrm{Var}_{t-1}(\vec{\alpha}_t)B' + \mathrm{I}_{x_b - x_a + 1} \cdot \sigma_\epsilon^2 \right)^{-1}$$

is the Kalman gain at time t.

The second step of the algorithm utilizes the Kalman smoother,

$$\begin{cases} \mathrm{E}_{t_b}[\vec{\alpha}_{t-1}] = \mathrm{E}_{t-1}[\vec{\alpha}_{t-1}] + J_{t-1}(\mathrm{E}_{t_b}[\vec{\alpha}_t] - \mathrm{E}_{t-1}[\vec{\alpha}_t]) \\ \mathrm{Var}_{t_b}(\vec{\alpha}_{t-1}) = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1}) + J_{t-1}(\mathrm{Var}_{t_b}(\vec{\alpha}_t) - \mathrm{Var}_{t-1}(\vec{\alpha}_t))J_t' \\ \mathrm{E}_{t_b}[\vec{\alpha}_t \vec{\alpha}_t'] = \mathrm{Var}_{t_b}(\vec{\alpha}_t) + \mathrm{E}_{t_b}[\vec{\alpha}_t](\mathrm{E}_{t_b}[\vec{\alpha}_t])' \end{cases},$$

and the lag-1 covariance smoother,

$$\begin{cases} \mathrm{Cov}_{t_b}(\vec{\alpha}_{t_b}, \vec{\alpha}_{t_b-1}) = (\mathrm{I}_m - \mathcal{K}_{t_b} B)A\mathrm{Var}_{t_b-1}(\vec{\alpha}_{t_b-1}) \\ \mathrm{Cov}_{t_b}(\vec{\alpha}_{t-1}, \vec{\alpha}_{t-2}) = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})J_{t-2}' + J_{t-1}(\mathrm{Cov}_{t_b}(\vec{\alpha}_t, \vec{\alpha}_{t-1}) - A\mathrm{Var}_{t-1}(\vec{\alpha}_{t-1}))J_{t-2}' \\ \mathrm{E}_{t_b}[\vec{\alpha}_t \vec{\alpha}_{t-1}'] = \mathrm{Cov}_{t_b}(\vec{\alpha}_t, \vec{\alpha}_{t-1}') + \mathrm{E}_{t_b}[\vec{\alpha}_t](\mathrm{E}_{t_b}[\vec{\alpha}_{t-1}])' \end{cases},$$

where $J_{t-1} = \mathrm{Var}_{t-1}(\vec{\alpha}_{t-1})A'(\mathrm{Var}_{t-1}(\vec{\alpha}_t))^{-1}$. By applying the two sets of equations above recursively for $t = t_b, t_b - 1, \ldots$, the required expectations can be obtained readily.

The update equation for σ_ϵ^2 contains two additional expectations, namely $\mathrm{E}[\vec{y}_t\vec{\alpha}_t']$ and $\mathrm{E}\left[\vec{y}_t'\vec{y}_t\right]$. These two expectations are computed as follows:

$$\mathrm{E}[\vec{y}_t\vec{y}_t'] = \vec{\tilde{y}}_t\vec{\tilde{y}}_t';$$
$$\mathrm{E}[\vec{y}_t\vec{\alpha}_t'] = \vec{\tilde{y}}_t\mathrm{E}_{t_b}[\vec{\alpha}_t]',$$

where $\vec{\tilde{y}}_t$ represents the realization of $\vec{y}_t$ and $\mathrm{E}_{t_b}[\vec{\alpha}_t]$ can be evaluated by the Kalman smoother.