Managing longevity risk

Li, Hong

Publication date:
2015

Document Version
Early version, also known as pre-print

Managing Longevity Risk

Hong Li

June 23th, 2015

Managing Longevity Risk

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op dinsdag 23 juni 2015 om 16.15 uur door

Hong Li

geboren op 8 april 1988 te Shantou stad, Guangdong Provincie, P.R. China.

Promotiecommissie:

Promotores:	prof. dr. Anja De Waegenaere
	prof. dr. Bertrand Melenberg
Overige Leden:	prof. dr. Hans Schumacher
	prof. dr. Mitja Stadje
	dr. Katrien Antonio
	dr. Ramon van den Akker

Acknowledgements

This thesis is the end of my journey in obtaining my Ph.D., it could not come to an end without the support and encouragement of numerous people during the journey. At the end of my thesis, I would like to thank all those people who made this thesis possible and an unforgettable experience for me.

The most important people in my Ph.D. study are my supervisors prof. Anja De Waegenaere and prof. Bertrand Melenberg, to whom I would like to express my deep gratitude, for their patience, motivation, enthusiasm, and immense knowledge. Bertrand was my master thesis supervisor in 2012. He brought me into the realm of longevity risk, and guided me through the application of the NWO project, "Managing Longevity Risk". I worked together with Anja and Bertrand during my Ph.D. period. I benefitted a lot from our weekly meetings. Their enthusiasm for econometrics, economics, and finance, in-depth thinking, strong work ethic, and diligence greatly influence the way I work. Their guidance helped me in all the time of research and writing of this thesis. Without their constant feedbacks and precious comments, this Ph.D. would not have been achievable.

I feel indebted to the committee members, prof. Hans Schumacher, prof. Mitja Stadje, dr. Katrien Antonio, and dr. Ramon van den Akker, for their time, interest, helpful comments, and detailed remarks to this thesis. I have also learnt a lot from inspiring discussions with Michel Vellekoop, Tim Boonen, Geng Niu, Ruixin Wang, and Yifan Yu. I appreciate useful comments from seminar participants at Tilburg University, University of Amsterdam, Wuhan University, Nanjing University, and Renmin University of China, and conference participants at Netspar Pension day 2013 and 2014, Longevity 9 and 10, and the 18th International Congress on Insurance: Mathematics and Economics.

I will never forget the suggestions and support offered during my stressful job search period. I was greatly supported by my supervisors. They gave me numerous useful comments on my job market paper as well as other application materials. I am grateful to

my references, Mitja Stadjes and Hans Reijnierse. I am also thankful for the useful advice from the Tilburg Ph.D. job search committee: Charles Noussair, Otilia Boldea, Bettina Drepper, Cecile de Bruijn, Patricio Dalton, and Burak Uras. Arthur Van Soest, Pavel Čížek, and Martin Salm also shared with me their helpful tips on my mock interview. I am also gratitude for the enormous help and encouragement from Yang Zhou, Ruixin Wang, Hao Liang, Geng Niu, Yifan Yu, and Lei Shu.

Many thanks to my many student colleagues for providing a stimulating and fun environment in which to learn and grow. My special thanks goes to my office mate Mitzi Perez Padilla. Sharing an office with you during these years was one of the pleasures of doing a Ph.D. Although we have different working schedules, we always have a lot to talk when we are together. And the staff from Department of Econometrics & OR, CentER and TiSEM, Anja, Ank, Bibi, Corine, Cecile, Carina, Heidi, Korine, Lenie, Nicole, Petra, Sandra and Yvonne, you made Tilburg University feel so much like a home, and thank you again for all your help.

Being surrounded by many friends made my Ph.D. life enjoyable in Tilburg University. Yang Zhou, Geng Niu, Jinghua Lei, Yun Wang, Ruixin Wang, Kun Zheng, Hao Liang, Ran Xing, Yifan Yu, Yi He, Hailong Bao, Yifan Zhang, Lei Shu, Bo Zhou, Yan Xu, Manxi Luo, Yuxin Yao, Chen sun, Yilong Xu, Zhuojiong Gan, Keyan Wang, Xu Lang, Di Gong, Yiyi Bai, Yufeng Huang, Bowen Luo, Chungyu Hung, Xue Jia, Zhengyu Li, Lu Yang, Miao Nie, Ying Yang, Huaxiang Yin, Yuejuan Yu, Zhiyu Yu, Baiquan Zhai, thank you all so much for your friendship.

I would like to express my sincere gratitude to my girlfriend Zhenzhen Fan, who has been by my side throughout this Ph.D., living every single minute of it, and without whom, I would not have had the courage to embark on this journey in the first place. You have given me a lot of inspiration and helpful comments on every single aspect of my research: research idea, model derivation, writing, and presentation skill. I benefitted a lot from your expertise in quantitative finance during the writing of this thesis. Moreover, my life is made much happier by you, and I could not have imagined what would it be if without your company for the past several years. No words can express my deepest love to you, and deepest thankfulness to your love, understanding and endless support.

Hong Li

May 7 2015

Contents

Chapter 1

Introduction

Life expectancy in most parts of the world has increased substantially in the past few decades. For example, the life expectancy at birth for a Dutch male increased from 70.3 years in 1950 to 79.9 in 2014, while this number for a Dutch female increased from 72.6 years in 1950 to 83.3 years in 2014.[1] Formally, we refer to longevity risk as the risk due to unexpected changes of the mortality rates of populations. During recent years, longevity risk has attracted more and more attention from pension plans and annuity providers. In particular, as an innovative way to transfer longevity risk, the longevity-linked capital market is developing rapidly. According to the records on Artemis,[2] the volume of transactions in the U.K. longevity-linked capital market has increased from £3.4 billion in year 2009 to £21.9 billion in year 2014. In this thesis we consider the management of longevity risk in the view of pension plans and annuity providers, including the forecast of future mortality rates and the hedging of mortality risk using longevity-linked derivatives.

Chapter 2 studies the forecast of future mortality rates. Most currently used mortality models are based on a linear extrapolation approach, meaning that the appropriately transformed quantities are assumed to follow a linear trend. For example, the Lee and Carter (1992) model and its variants are heavily used by actuarial and statistics agencies, such as the Actuarial Society and Statistics Netherlands. However, many mortality data sets do not show clear linear trends. For example, it is shown in Chapter 2 that, the remaining life expectancy of U.S. and Dutch populations generated by the Lee and Carter (1992) model, one of the most used mortality models in the past two decades, is rather sensitive to the chosen sample size. Therefore, an important question to ask when applying a linear extrapolation model is: What sample size should one use? The choice of sample size should not only depend on the in-sample fit, but also the out-of-sample forecast accuracy. Moreover, the choice of sample size might also depend on the

[1] Source: Statistics Netherlands, http://statline.cbs.nl/.

[2] http://www.artemis.bm/library/longevity_swaps_risk_transfers.html.

underlying population, as the mortality pattern may be different among populations. In Chapter 2 we try to answer the above question. In particular, we propose a Bayesian learning approach to determine the choice of (the posterior distribution of) the sample size. The posterior distribution of the sample size is allowed to be age-, gender-, forecast-horizon-, model-, and country-specific. Using simulation studies, we find that the Bayesian learning method is able to cope with not only linear structural breaks, but also accelerating trends in the mortality data. Finally, we apply the Bayesian approach to U.S. and Dutch data using the Lee and Carter (1992) and the Cairns et al. (2006) model, and see that clear suggestions of the choice of sample size is found for most of the age-, gender-, forecast-horizon-, and model combinations. In the out-of-sample forecast analysis, we see that the Bayesian model outperforms the original models with the minimal DIC ratio in the majority of scenarios.

Given the uncertainty regarding the distribution of future mortality rates, it is important for pension plans and annuity providers to have a risk management strategy that is robust with respect to possible estimation inaccuracies of the probability distribution of the mortality rates. In Chapter 3 we consider a static setting where the insurer is not certain about the probability distribution of the future mortality rates, but optimizes against the worst-case objective function (mean-variance or mean-conditional-value-at-risk). The worst-case is with respect to probability distributions within a compact set characterized by the Kullback-Leibler divergence. We choose the parameters characterizing the compact set in the way such that it can be treated as a confidence set around the best estimated probability distribution of the future mortalities. The robust optimization approach can be used in combination with the Bayesian learning approach studied in Chapter 2. The Bayesian learning approach can be used to gain more accurately forecasts when we want to fit a particular mortality model to a particular data set. However, even though one applies the Bayesian approach, the resulting nominal distribution of the future mortality rates might still be subject to estimation inaccuracy. Therefore, it still makes sense to apply the robust optimization approach. Moreover, the robust optimization approach is able to incorporate more general model risks, such as uncertainties regarding the parametric form of the model.

As the longevity-linked capital market is at its early stage, the types of longevity-linked derivatives are rather limited. In particular, it is difficult to buy products with

payments contingent on any desired cohorts and with long maturities. Therefore, except for the population basis risk, i.e., the mismatch of mortality experience of the portfolio-specific population and the national population, the insurer would also be concerned about cohort mismatch and maturity mismatch. In Chapter 3, we apply the robust optimization method to Dutch national data and a Dutch collective pension portfolio. In the application, we assume that the insurer is committed to annuity payments of multiple cohorts, where she can only trade the survivor swap which has the same maturity as the liabilities, but is contingent on one cohort. In other words, we consider as an illustration the hedge in the presence of only population basis risk and cohort mismatch. However, the model proposed in Chapter 3 is applicable to more general situations, e.g., we can allow for hedging instruments with payments contingent on any desired cohorts and with any maturities. In Chapter 4, we take into account the liquidity with respect to longevity-linked derivatives with long maturities, and we explicitly incorporate hedging instruments with different maturities in the framework. In the application in Chapter 3, we show that the robust optimization performs on average better than the non-robust optimization when moderate degrees of estimation inaccuracy exist.

Chapter 3 considers the management of longevity risk in a static setting. Although static hedging is nowadays commonly used in longevity risk hedging, it would be helpful to explore dynamic hedging possibilities, as is done in Chapter 4. In particular, we consider in this chapter a dynamic value hedging problem: The hedger wishes to minimize the variance of her hedging error, defined as the deviation of the market value of her investments to the market value of her liabilities, valued at some future time, T. The variance criterion is commonly used by researchers and practitioners in static settings. However, in the dynamic setting, the variance criterion might be also easier to interpret than utility functions. For example, we can measure the hedging quality by looking at the optimal variance of the hedging error, without resorting to the functional form of the utility functions and the choice of risk aversion parameters. The latter is not straightforward to specify for a pension plan. Moreover, the use of the variance criterion is relevant for hedging longevity risk. An insurance company which sells both annuities and life insurance products faces longevity risk in two directions. The company loses profits from the life insurance products in case of unexpected mortality deteriorations, e.g., catastrophe events, while it loses profits from the annuities when the life expectancy of

their annuitants turn out to be higher than expected. Therefore, the insurance company would like to minimize the deviation of its hedging error from both directions. A similar value hedge setting is considered in Cairns (2013) and Cairns et al. (2014) in a static framework.

An important property of dynamic hedging strategy is time-consistency. A time-consistent strategy is a strategy that will be followed by the hedger at any time and in any state of the world. Two types of time-consistent hedger are studied in literature: naive time-consistent hedger and sophisticated time-consistent hedger (Grenadier and Wang 2007). In particular, a naive time-consistent hedger places a strategy at the initial date, and commits herself not to deviate from the initial strategy at later dates, even when she has the incentive to do so. In contrast, a sophisticated time-consistent hedger places a strategy at the initial date, taking into account the possibilities of deviating from the initial strategy at later dates. Therefore, the optimal strategy chosen by the sophisticated time-consistent hedger is the one that will be followed at later dates when she re-optimizes using the same objective function. In Chapter 4, we consider the optimal hedging strategy of a sophisticated time-consistent hedger under the variance criterion.

Despite the rapid growth, the longevity-linked capital market is still at its early stage of development. As a result, the liquidity of longevity-linked derivatives is currently rather limited. To cope with this fact, we derive optimal hedging strategies not only in the case where all assets can be traded continuously, but also in the case where part of the assets, in particular the longevity-linked derivatives, can only be traded at a deterministic and lower frequency.

To sum up, this thesis provides findings for forecasting future mortality rates, and for hedging longevity risk for pension plans, annuity providers, and insurance companies. The mortality forecasting and the longevity risk hedging are both fast-developing fields, and there are many possible future research directions along the findings in this thesis. Interesting directions include the construction of mortality models with more robust estimations and forecasts with respect to the sample size, especially in the multi-population context; the robust dynamic longevity risk management; and the dynamic longevity risk hedging with illiquid assets in a richer setting, for example, with stochastic trading frequency and transaction costs, etc.

Chapter 2

The Choice of Sample Size for Mortality Forecasting: A Bayesian Learning Approach[3]

2.1. Introduction

Compared with the extensive literature on mortality forecasting, relatively little is known about the choice of sample size for the existing mortality models. Most mortality models proposed in the past twenty years are based on a "linear extrapolation" approach, meaning that the evolution of the modeled quantities (such as the log of central death rates or the logit[4] of death probabilities) is assumed to follow a linear trend. Two most widely used models in this class are the ones proposed by Lee and Carter (1992) (the Lee-Carter model) and Cairns, Blake, and Dowd (2006) (the CBD model).[5] Various extensions to these two models have been proposed, including Brouhns et al. (2002), Li and Lee (2005), Renshaw and Haberman (2006), to name a few. See Cairns et al. (2009), who summarize several most popular models in the linear extrapolation class.

An immediate result of using linear extrapolation models is that the forecasted mortality rates are highly sensitive to the choice of sample size. This phenomenon has been documented by many authors. In their original paper, Lee and Carter (1992) estimate the model based on U.S log mortality data from 1900 to 1989. They argue that the length of the sample size is not critical as long as it is longer than 20 years. However, Lee and Miller (2001) restrict the sample size to start from 1950, and a better fit is obtained.

[3]This chapter is coauthored with Anja De Waegenaere and Bertrand Melenberg.

[4]logit $x = \log \frac{x}{1-x}$ for $x \in (0, 1)$.

[5]Besides the linear extrapolation models, there are other classes of mortality models, see, for example, Börger (2009), Börger et al. (2013), Miltersen and Persson (2005), Bauer et al. (2010), Plat (2011), and Börger (2010). However, these models have more complex structures and are not considered in this chapter.

Baran et al. (2007) forecast the mortality rate of the Hungarian population from 2004 to 2040 using the Lee-Carter model based on mortality data from 1949 to 2003. As a result, the forecasted mortality rates for males from age 45 to 55 are increasing, which does not seem to be reasonable. However, when they apply the Lee-Carter model with data from 1989 to 2003, both forecasted mortality for males and females are decreasing. Cairns et al. (2006) apply the CBD model to the England & Wales male data, and get significantly different forecasts when using data from 1961 to 2002 than when using data from 1981 to 2002. See Pitacco et al. (2009) and van Berkum et al. (2013) for more thorough summaries.

In this chapter, we aim to answer the following question: If one wishes to fit some linear mortality model to some data set, what sample size should one choose? To answer this question, we propose a Bayesian learning approach to determine model-specific posterior distributions of the sample size for a class of linear mortality models.[6] In particular, the sample size is included as an extra parameter into the parameter space of the mortality model and its (conditional) posterior distribution is updated based on the historical performance for different forecast horizons. Briefly speaking, we first choose a set of sample sizes, estimate the mortality model based on each of the sample sizes, then perform the out-of-sample forecast for different forecast horizons using each model. A sample size receives a larger posterior probability mass if the corresponding model yields a larger likelihood in the out-of-sample forecast; this can be seen as a weighted average of the in-sample fit and the out-of-sample forecast performance.

First of all, this chapter contributes to the literature on sample size determination for mortality forecasting. Despite its importance, only few studies focus on this topic. Booth et al. (2002) provide tests for the optimal sample size of the Lee-Carter model. Denuit and Goderniaux (2005) model the time effect in the Lee-Carter model as a linear function of time, and look for a sample size such that the adjustment coefficient R^2 is maximized. O'Hare and Li (2012) and van Berkum et al. (2013) study structural changes in mortality rates using several mortality models. In these studies, the authors first estimate a particular model, and then detect structural breaks in the implied latent

[6]There is an extensive literature on the topic of Bayesian sample size determination, such as Adcock (1997), De Santis (2004), Wang and Gelfand (2002), and Weiss (1997), to name a few. However, the focus of these studies is different from our focus, including, for example, determining the minimal sample size such that a hypothesis test has enough power or a desired estimation precision can be achieved.

time process. A common feature of the existing literature is that it looks for a sample size for which the implied time effect process is close enough to linear so that the underlying model specification is considered to be valid.

Our method has several advantages compared with the existing studies. First, our method allows for an age-specific posterior distribution of the sample size, while using data of all ages in the estimation. All the aforementioned studies determine the optimal sample size by focusing at the underlying latent time process. Since this time process is an aggregation of the time effects of the different ages, no age-specific information can be elicited from it.In our model, age- and gender-specific posterior distributions of sample sizes can be naturally computed, while data for both genders and multiple ages are used in the estimation and learning process. Second, we look at the choice of sample size in a probabilistic way, and thus avoid sharp transitions in the use of data. In O'Hare and Li (2012) and van Berkum et al. (2013), structural breaks are allowed to exist. If such a structural break is detected, then only data after this break should be used and data before the new break will be discarded. In contrast, our method results in a smoother transition in the use of data when facing a structural break (or other nonlinearities): When using a Bayesian learning method, a break gains more and more posterior weight only when more and more subsequent observations support its existence. (We provide more concrete illustrations via simulation in Section 5.) Third, by choosing different out-of-sample forecast horizons in the updating process, we may gain better insights in the suitability of different mortality models. Insensitivity of the posterior distribution of the sample size to the forecast horizon can be seen as an indicator that the mortality model is suitable as a forecast model: The selected sample seems to support the linearity assumption of the model. However, sensitivity of the posterior distribution of the sample size to the forecast horizon suggests that the linearity assumptions does not fit the data well, so that the mortality model is not likely to be suitable as a forecast model.

This chapter also contributes to the literature of Bayesian mortality modeling. As will be discussed later, our method consists of two parts: Estimating the underlying mortality model for a fixed sample size, and estimating the posterior distribution of the sample size. There are many existing studies focusing on the first part, i.e., the Bayesian analysis of mortality models for a fixed sample size. For example, Pedroza (2002), Pedroza (2006), Czado et al. (2005), Reichmuth and Sarferaz (2008), Kogure et al. (2009), and

Kogure and Kurachi (2010) study the Bayesian modelling of the Lee-Carter model and its several extensions. Moreover, Cairns et al. (2011) propose a Bayesian estimation method for a two population mortality model. For the Bayesian mortality modelling, given a fixed sample size, our method extends the state space formulation proposed by Pedroza (2002) and Pedroza (2006) to a more general setup. Also, correlations among multiple populations are explicitly allowed for. Our formulation is applicable to various existing linear mortality models, such as most model specifications studied in Cairns et al. (2009). (More details are discussed in Section 4.)

Although our method has a wide applicability, we mainly focus on the original Lee-Carter and the CBD model with the underlying time effect process modeled by a random walk with drift. The reason is that the simplicity of these model specifications allows us to demonstrate our approach in a clearer way. We apply our method to U.S. and Dutch data, and find quite concentrated posterior distributions for most combinations of age, gender, forecast horizon, and country for both mortality models, where these posterior distributions turn out to be age- and gender-specific. We also compare the out-of-sample forecast performance of the Bayesian models and the original models with the optimal Deviance Information Criterion (Spiegelhalter et al. 2002) ratios for the Dutch data. In particular, the mean squared errors of the out-of-sample forecasts from 2000 to 2009 are compared. We find that the Bayesian models outperform the original models in the majority of the cases.

The remainder of the chapter is organized as follows. In the next section a brief introduction to the Lee-Carter model and the CBD model is provided. In Section 3 we discuss the importance of sample size determination in more details. In Section 4 a formal description of the Bayesian model is given. In Section 5 we evaluate the performance of our method with simulated data. Section 6 presents the empirical results. Key results of the out-of-sample forecast is reported in Section 7. Section 8 presents the conclusion. The appendix contains some technical details. Complementary results are presented in an online appendix.

2.2. Mortality models

In this section we introduce the notations used in this chapter and give a brief description of the Lee-Carter (Lee and Carter 1992) and the CBD model (Cairns et al. 2006).

2.2.1. Notation

We construct an index, i, belonging to an index set I, with $i = (x, t, g)$, where x, t, and $g \in \{m, f\}$ denote age, time, and gender, respectively. In other words, the index $i = (x, t, g)$ represents the group of individuals aged x at time t with gender g. For any $i = (x, t, g)$:[7]

- m_i is the crude central death rate. Specifically, $m_i = \frac{D_i}{E_i}$, where D_i is the observed number of deaths for i, and E_i the corresponding exposure.

- q_i is the one year death probability, which can be approximated by the crude death rates via (see, for example, Pitacco et al. 2009)

$$q_i \approx 1 - \exp(-m_i). \tag{2.1}$$

- $\mathcal{M}$ is the set of the stochastic mortality models considered.

- $\mathcal{J} \subset \mathbb{N}$ is the set of sample sizes. For any $J \in \mathcal{J}$, J is the length of the sample period upon which we estimate a model. For example, if the end year is 1999, then $J = 30$ means mortality data from year 1970 to 1999, and $J = 20$ means mortality data from year 1980 to 1999. We denote $J_{min} = \min\{J \in \mathcal{J}\}$ and $J_{max} = \max\{J \in \mathcal{J}\}$.

- Θ_M is the set of parameters for the stochastic mortality model $M \in \mathcal{M}$. Moreover, we define $\Theta_M = \tilde{\Theta}_M \times \mathcal{J}$, where $\tilde{\Theta}_M$ is the space of parameters except for the sample size. The structure of Θ_M is model-specific, i.e., its composition varies with the stochastic mortality model. In the Appendix, we specify the parameter sets for the Lee-Carter model and for the CBD model.

[7] We use notations such as $q_{x,t,g}$ and q_i interchangeably, according to the specific context. There is no difference in their meaning.

2.2.2. The stochastic mortality models

We focus on two most widely used stochastic mortality models: the Lee-Carter model (Lee and Carter 1992) and the CBD model (Cairns et al. 2006). Therefore, we have $\mathcal{M} = \{LC, CBD\}$. In particular, we restrict ourselves to the case where the underlying latent time process follows a random walk with drift.[8] For each model, we fit the model to both the male and female population, where correlation between these two populations are explicitly modeled. A brief description of the two models is provided next.

Lee-Carter model

The Lee-Carter model postulates that, for any $i = (x, t, g)$, the central death rate m_i is given by

$$\ln m_i = \alpha_{x,g} + \beta_{x,g}\kappa_{t,g} + \varepsilon_i,\ \varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2_{x,g}), \tag{2.2}$$

where ε_i is the corresponding error term, and $\mathcal{N}(0, \sigma^2_{x,g})$ is the normal distribution with mean 0 and variance $\sigma^2_{x,g}$. We model the two time varying parameters $\boldsymbol{\kappa}_t = [\kappa_{t,m}, \kappa_{t,f}]'$ as a random walk with drift process

$$\boldsymbol{\kappa}_t = \boldsymbol{d} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_\omega), \tag{2.3}$$

where $\boldsymbol{d} = [d_m, d_f]'$, $\boldsymbol{\omega}_t = [\omega_{t,m}, \omega_{t,f}]'$, and Σ_ω is the covariance matrix of $\boldsymbol{\omega}_t$. In order to estimate the model, some normalization has to be imposed. In this chapter, we make the same normalization as in the original paper:

$$\sum_{x \in N} \beta_{x,g} = 1, \text{ and } \sum_{t \in T_J} \kappa_{t,g} = 0,$$

for $g \in \{m, f\}$, where N and T_J is the set of ages and years, respectively, where the subscript J indicates the dependence of the time set on the underlying sample size.

[8] This restriction is mainly for expositional purposes. In Section 4 we discuss the class of models to which the Bayesian sample size selection method can be applied.

CBD model

For any $i = (x, t, g)$, the CBD model (Cairns et al. 2006) postulates that the logit of the one year death probability q_i is given by

$$\text{logit } q_i = \log(\frac{q_i}{1-q_i}) = \kappa_{t,g}^{(1)} + \kappa_{t,g}^{(2)} x + \varepsilon_i,\ \varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma_{x,g}^2), \tag{2.4}$$

where we assume that the time varying parameters $\boldsymbol{\kappa}_t = [\kappa_{t,m}^{(1)}, \kappa_{t,m}^{(2)}, \kappa_{t,f}^{(1)}, \kappa_{t,f}^{(2)}]'$ follow a random walk with drift process

$$\boldsymbol{\kappa}_t = \boldsymbol{d} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{\omega}_t,\ \boldsymbol{\omega}_t \overset{i.i.d}{\sim} \mathcal{N}(0, \Sigma_\omega), \tag{2.5}$$

where $\boldsymbol{d} = [d_{1,m}, d_{2,m}, d_{1,f}, d_{2,f}]'$, $\boldsymbol{\omega}_t = [\omega_{m,t}^{(1)}, \omega_{m,t}^{(2)}, \omega_{f,t}^{(1)}, \omega_{f,t}^{(2)}]'$, and Σ_ω is the covariance matrix of $\boldsymbol{\omega}_t$.

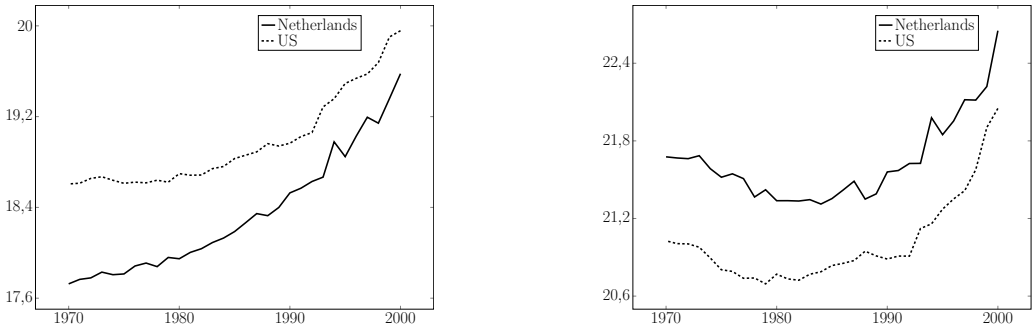
2.3. Sample size selection

If the mortality trend is linear, the choice of sample size is simple: We should use the largest sample size we can (from a statistical point of view). However, when the mortality trend is not linear, the problem becomes more complicated. For example, if the mortality data consists of two clear linear segments with different slopes, i.e., there is an obvious structural break, then we should probably only use the data after the break. However, if the mortality trend is, for example, accelerating, or there seems to be a structural break near the end of the sample, i.e., only few observations are available after the suspectable break, then it is not obvious which sample size one should use. In such cases, it is not clear, compared with the more recent data, how relevant the older data is to forecasting the future. A sample size yielding better in-sample fit does not necessarily produce better forecasts.

2.3.1. Importance of sample size

In this subsection, we illustrate the potential importance of the choice of sample size. Figure 2.1 displays the best estimates of the expected remaining lifetime for the U.S. and

the Dutch population of age 65 in 2010 and 2009, respectively.[9] The estimates are based on the Lee-Carter model using different sample sizes. In the figure we plot the starting year of the sample on the horizontal axis. In each case, the end year of the sample is $T = 2010$ for U.S. data, and $T = 2009$ for Dutch data. Hence, the corresponding sample size J equals $J = T -$ *starting year* $+1$. We can see that the estimated expected remaining lifetimes for both genders in both countries increase significantly as the sample size decreases. In particular, as the sample size decreases from 40 to 10 years, the estimated expected remaining lifetime increases by about 7.2% and 6.7% for the U.S. and the Dutch male population, and 4.7% and 2.7% for the female population, respectively. We see that the forecasts of the Lee-Carter model are sensitive to the underlying sample size. Therefore, the role of sample size determination is not negligible.



(a) Male (b) Female

Figure 2.1: The expected remaining lifetime for the U.S. and the Dutch population of age 65 in 2010 and 2009, respectively. The left figure reports the result for the male population and the right figure for the female population. The calculations are based on the Lee-Carter model using different sample sizes. The x-axis denotes the *starting year* of the sample. The end year for the U.S. and the Dutch population are 2010 and 2009, respectively.

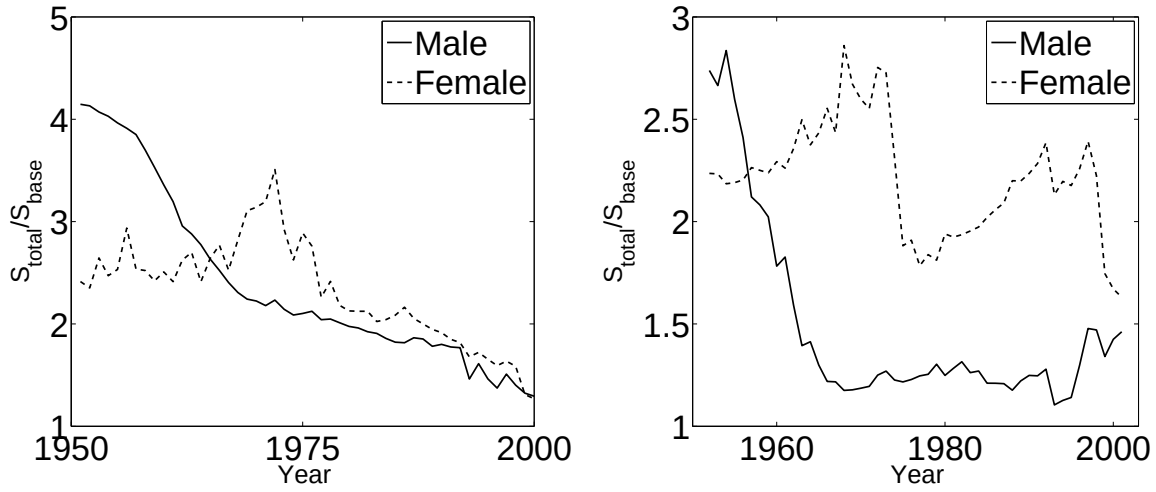
2.3.2. Existing methods to select a sample size

Booth et al. (2002) propose a method to find the optimal sample size for the Lee-Carter model, where the optimal sample size identifies the largest sample size for which the estimated κ introduced in Section 2 is reasonably close to a linear process. In particular, the authors first compute for each sample size a measure to quantify the total lack of fit caused by treating the κ process as a fitted linear process, S_{total}, together with the base

[9]The end years we use are the most recent years for which data is available for these two countries. The best estimates of the expected remaining lifetime are obtained by extrapolating the $\boldsymbol{\kappa}$ process, setting the error terms equal to zero.

lack of fit caused by using the estimated κ-s, S_{base}. Since S_{total} includes additional lack of fit compared to S_{base}, the ratio S_{total}/S_{base} will be larger than or equal to one for all sample sizes, and the closer this ratio is to one, the closer the κ process is to linear.

Booth et al. (2002) apply their method to Australian data, and find a clear choice of starting year at around 1968. We apply their method to Dutch and U.S. data. The results are reported in Figure 2.2. From the figure, we see a sharp drop of the ratio at 1965 for the U.S. male. However, for U.S. females and both Dutch males and females, the choice of the sample size is not obvious. For the U.S. female, a sharp drop happens at 1974, but the ratio is not stable afterwards. For the Dutch population, the ratios are increasing with the sample size, indicating that the decrease of the mortality rate is probably accelerating for both genders. Therefore, for the U.S. female and the Dutch data, clear conclusions regarding the choice of sample size cannot be drawn from the method of Booth et al. (2002), and other methods of determining the sample size seem to be needed.



(a) Netherlands (b) US

Figure 2.2: The application of the method in Booth et al. (2002) to the Dutch and the US data. Data from age 21 to 89 and year 1950 to 2009 is used for the Netherlands, and from year 1950 to 2010 is used for the U.S.. The horizontal axis denotes the *starting year* of the sample.

An alternative approach would be to choose the sample size that yields the minimal DIC (Deviance information criterion) ratio. DIC is a Bayesian version of the AIC (Akaike information criterion) and BIC (Bayesian information criterion), see, e.g., Spiegelhalter et al. (2002). Similar to AIC and BIC, it offers a relative estimate of the information lost

when a given model is used to represent the process that generates the data. As such, it provides a means for model selection. The details of this approach are explained in the Appendix. In Section 7, we will compare the out-of-sample forecast accuracy when sample size is determined based on DIC, to the out-of-sample forecast accuracy resulting from the Bayesian sample selection method that we propose in this chapter.

2.4. The Bayesian model

In this section, we give a brief introduction to the Bayesian learning approach that we propose. Throughout the remainder of the chapter, we will use the notation $[t_1 : t_2] = \{t_1, t_1 + 1, \cdots, t_2 - 1, t_2\}$, for $t_1 \leq t_2 \in \mathbb{N}$.

Let mortality model $M \in \mathcal{M}$ be given, and suppose that we have at our disposal a set of mortality data indexed by $[T_0 : T]$, i.e., T_0 is the starting year of our sample, and T is the last year for which data is available. For each $t \in [T_0 : T]$, we denote by y_t the mortality quantities at time t.[10] Moreover, we let Y_t be the observed mortality quantities up to time t, i.e., $Y_t \equiv \{y_{T_0}, y_{T_0+1}, ..., y_t\} = y_{[T_0:t]}$.

The method that we propose allows to determine age- and gender-specific posterior distributions of the sample size. In such cases, y_t equals the mortality quantities for specific age- and gender-groups only.

2.4.1. The basic idea

The basic idea of the Bayesian learning approach is that part of the latest in-sample data is reserved for Bayesian learning, $y_{[\widetilde{T}:T]}$ for $\widetilde{T} \leq T$. We first choose a set of sample size, and update the posterior distribution of the sample size based on the likelihood of the data used for learning. For any given mortality model $M \in \mathcal{M}$, the procedure is as follows:

1. Choose:

   (a) a set of sample sizes, $\mathcal{J}$,[11]

[10] The mortality quantities will be model dependent: y_t includes the log of the central death rates for the Lee-Carter model and the logit of the one year death probabilities for the CBD model.

[11] The set of sample sizes can in principle depend on the underlying mortality model. However, for the purpose of comparison, we impose the same set of sample sizes for each mortality model and for the same data set.

(b) a prior distribution over the set $\mathcal{J}$,

(c) a forecast horizon $\mu \in \mathbb{N}$,

(d) a value for $\widetilde{T} \in [T_0 : T]$; $\widetilde{T}$ is the first year in which the prior distributions of the model parameters and the sample size will be updated.

These choices should satisfy

$$J_{max} + \mu \leq \widetilde{T} - T_0 + 1. \tag{2.6}$$

Condition (2.6) ensures that for all sample sizes in the set $\mathcal{J}$, both the data that is needed to estimate the model, and the data that is needed for the Bayesian updating of the prior distributions in year $\widetilde{T}$ using a forecast horizon of μ years, are included in the sample period $[T_0 : \widetilde{T}]$.

2. Apply the Bayesian learning approach. This learning approach is illustrated in Figure 2.3.

   (a) For each $J \in \mathcal{J}$, estimate (the posterior distribution of the parameters of) the model in year $\widetilde{T}$, based on the sample $y_{[\widetilde{T}-\mu-J+1:\widetilde{T}-\mu]}$. See line (a) in Figure 2.3.

   (b) Derive the posterior predictive distribution of $y_{\widetilde{T}}$, and update the posterior distribution of $J \in \mathcal{J}$, based on the corresponding likelihood with respect to $y_{\widetilde{T}}$.

   (c) Forward by one period. Estimate the model based on the sample $y_{[\widetilde{T}-\mu-J+2:\widetilde{T}-\mu+1]}$ for each $J \in \mathcal{J}$, and update the posterior distribution of J in year $\widetilde{T}+1$, based on the corresponding likelihood of $y_{\widetilde{T}+1}$. See line (b) in Figure 2.3.

   (d) Continue the process until the final step, where the update is based on the likelihood of y_T. See line (c) in Figure 2.3.

To illustrate the procedure, we give a concrete example. Let $[\tilde{T} : T] = [1991 : 2010]$, i.e., consider the Bayesian learning approach where a period of 20 years is used for the updating procedure. Also, let the forecast horizon be $\mu = 20$ and let the set of sample sizes be $\mathcal{J} = \{10, 11, \cdots, 39\}$. Hence, the updating process starts at $\tilde{T} = 1991$. To

derive the posterior predictive distribution of y_{1991}, we estimate the model based on the sample $y_{[1962:1971]}$ for $J = 10$, based on the sample $y_{[1961:1971]}$ for $J = 11$, and so on up to $J = 39$ based on the sample $y_{[1933:1971]}$. We use these estimates to derive the posterior predictive distribution of $y_{1971+\mu} = y_{1991}$, and to update the posterior distribution of $J \in \mathcal{J}$. We then repeat this procedure for the cases where the end year of the sample period used for estimating the model is 1972, 1973, and so on until 1991 in the final step.

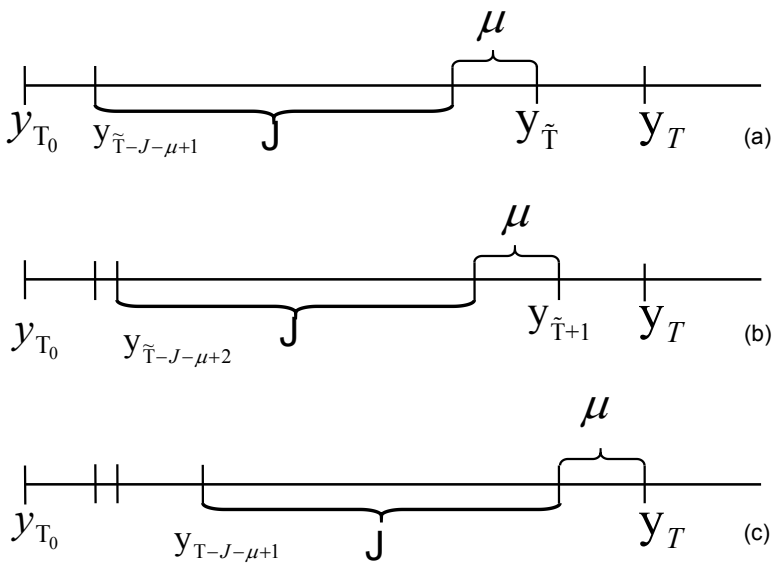


Figure 2.3: A brief illustration of the updating procedure of the method.

To sum up, for any given model and forecast horizon, μ, we update the posterior distribution of the sample size at each step based on the likelihood of the observed data μ years ahead. Our method can be performed using only one μ, but the use of multiple μ-s can give extra flexibility. For example, the posterior distribution for a smaller μ can be used if one is interested in short-term forecasting, and the posterior distribution for a larger μ can be used if one is interested in long-term forecasting. Alternatively, the use of different μ-s can serve as a diagnostic test. If for a model M the posterior distributions of the sample size turn out to be insensitive to different μ-s, this might be seen as evidence supporting the use of model M as a forecast model. In contrast, if the posterior distributions are sensitive to the change of μ, namely, the model seems to be too sensitive to the nonlinearities in the data, then it might not be suitable to be used as a forecast model.

2.4.2. The posterior sample size distribution

In this section, we discuss how, starting from a prior distribution for the sample size, the posterior distribution can be determined in each step of the Bayesian learning process described in the previous subsection.

Recall that $\widetilde{T}$ is the first year in which the probability distribution of the parameters θ_M is updated. Formally, for each $M \in \mathcal{M}$, we can write the joint posterior distribution of the parameters $\theta_M = (\tilde{\theta}_M, J)$ in year $t \geq \widetilde{T}$, given any age x, gender g, and forecast horizon μ as

$$p(\tilde{\theta}_M, J|Y_t, x, g, \mu, M) = p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M) \cdot p(J|Y_t, x, g, \mu, M). \tag{2.7}$$

In particular, our primary interest is in $p(J|Y_t, x, g, \mu, M)$. This posterior density is estimated by using an iterative updating algorithm. First, for every x, g, and μ, we specify the prior distribution $p(J|x, g, \mu, M)$. Suppose that we are at time t and we know $p(J|Y_{t-1}, x, g, \mu, M)$. Given J**,** we estimate the conditional posterior distribution of $\tilde{\theta}_M$ using data $y_{[t-\mu-J+1:t-\mu]}$. This posterior distribution is denoted by $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M)$. Then, we update $p(J|Y_{t-1}, x, g, \mu, M)$ to $p(J|Y_t, x, g, \mu, M)$ using Bayes' formula

$$p(J|Y_t, x, g, \mu, M) \propto p(y_t|Y_{t-1}, J, x, g, \mu, M) \cdot p(J|Y_{t-1}, x, g, \mu, M), \tag{2.8}$$

for each J, where

$$\begin{aligned} &p(y_t|Y_{t-1}, J, x, g, \mu, M) \\ &= \mathbb{E}_{\tilde{\theta}_M}[p(y_t|\tilde{\theta}_M, Y_{t-1}, J, x, g, \mu, M)] \\ &= \int_{\tilde{\Theta}_M} p(y_t|\tilde{\theta}_M, Y_{t-1}, J, x, g, \mu, M) \cdot p(\tilde{\theta}_M|Y_{t-1}, J, x, g, \mu, M) d\tilde{\theta}_M. \end{aligned} \tag{2.9}$$

We continue the updating procedure until we obtain $p(J|Y_T, x, g, \mu, M)$. At every $t \geq \widetilde{T} + 1$, we treat the posterior distribution at $t-1$ as the corresponding prior distribution at t for every J, x, g, μ, and M. For the first year in which the density is updated, i.e., for $t = \widetilde{T}$, $p(J|Y_{t-1}, x, g, \mu, M)$ in (2.8) is replaced by the prior distribution $p(J|x, g, \mu, M)$.

To sum up, at every time $t \geq \widetilde{T}$, we first estimate the conditional posterior distributions $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M)$, and use them to update the corresponding posterior

distribution of the sample size. Because we have no knowledge about the weights of the sample size ex ante, we let the prior distribution be uniform in most of our analyses. In Subsection 6.2, we show that the posterior distributions of sample size are relatively insensitive to the choice of prior distribution.

Estimation of $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M)$

Because $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M)$ is the posterior distribution of the rest parameters given a sample size, standard techniques of Bayesian dynamic modeling can be used. In this section we propose a general state space formulation which incorporates most existing linear mortality models. This is the class so-called Time Varying-Vector Auto Regressive (TV-VAR) models, see, for example, Damien et al. (2013) for a more detailed overview.

Formally, suppose we have at our disposal a set of mortality quantities, $y_{[T_0:T]}$, where y_t is a $2n \times 1$ vector for each t (containing n ages for both gender), and we would like to estimate (the posterior distribution of) $\tilde{\theta}_M$ in year $\tilde{T}$. Given J and μ, we use data $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$ in the estimation. Furthermore, assume that the y_t process can be modeled as

$$y_t = B_t X_t + F_t' z_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \tag{2.10}$$

$$z_t = \sum_{i=1:p} G_{t,i} z_{t-i} + \omega_t, \quad \omega_t \sim N(0, \Sigma_{\omega_t}), \tag{2.11}$$

where, moreover, ε_t and ω_t are allowed to be correlated with each other for every $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, but where ε_t and ω_t are assumed to be independent of ε_s and ω_s with $s \in [\tilde{T} - \mu - J + 1 : t - 1]$. Model (2.10) – (2.11) includes, for example, all model specifications studied in Cairns et al. (2011) except for M2A and M3B.[12] In model (2.10) – (2.11), the parameter vector $\tilde{\theta}_M$ is a vector consisting of the components of B_t, X_t, F_t, $[G_{t,i}]_{i=1:p}$, Σ_{ω_t}, z_t, and Σ_{ε_t}, for $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$. Specifically, X_t is often a time-invariant matrix, such as the α_x-s in the Lee-Carter model; z_t is a vector representing the period effects and the cohort effects; B_t, F_t, and G_t are the corresponding parameters that could either be matrices or vectors, and are often time-invariant. Equation (2.11) is the vector-latent autoregressive process for the period effects and the cohort effects. The dimensions of X_t, z_t, B_t, F_t, and $\sum_{i=1:p} G_{t,i}$ depend on the

[12]The specifications M2A and M3A in Cairns et al. (2011) include a moving averaging part in the cohort effect process, and are thus not included in the TV-VAR formulation.

underlying model specifications.

Take the Lee-Carter model for example: in this case $p = 1$, $B_t = I_n$ for all t, $X_t = [a_{1,m}, a_{2,m}, ..., a_{n,m}, a_{1,f}, a_{2,f}, ..., a_{n,f}]'$ for all t,

$$F_t = \begin{bmatrix} b_{1,m} & b_{2,m} & ... & b_{n,m} & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & b_{1,f} & b_{2,f} & ... & b_{n,f} \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

for all t, $z_t = [\kappa_{t,m}, \kappa_{t,f}, d_m, d_f]'$,

$$G_{t,1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

for all t, and $\omega_t = [\omega_{t,m}, \omega_{t,f}]'$. Also, ε_t is the vector of error terms, Σ_ε is a diagonal matrix, and $\Sigma_{\omega_t} = \Sigma_\omega$ for all t.

In general, model (2.10) – (2.11) can be estimated using numerical algorithms. The estimation can be done in two steps. First, the posterior distributions of z_t at each time t can be obtained by backward filtering, see West and Harrison (1997) and Damien et al. (2013). Second, we can specify the conditional posterior distributions of other parameters. For example, one can specify the posterior distribution of Σ_ε (conditional on the other parameters) to be a Wishart distribution, etc. After all conditional posterior distributions of these other parameters are specified, we can estimate them, together with the posterior distribution of z_t-s, by the Gibbs sampler. This estimation algorithm is applicable to most existing mortality models. An example can be found in Pedroza (2006), who fits model (2.10) – (2.11) to a single gender Lee-Carter model. A detailed sampling procedure of the models used in this chapter can be found in the Appendix.

Estimation of $p(J|Y_t, x, g, \mu, M)$

Suppose that we have already estimated model (2.10) – (2.11), and have N draws of the parameters $\tilde{\theta}_M$ for each $J \in \mathcal{J}$. Denote the sample analogue of $p(y_t|Y_{t-1}, J, x, g, \mu, M)$

in Equation (2.9) by $p^N(y_t|Y_{t-1}, J, x, g, \mu, M)$. Then we have

$$p^N(y_t|Y_{t-1}, J, x, g, \mu, M) \propto \frac{1}{N}\sum_{\ell=1}^{N} p(y_t|\tilde{\theta}_M^{(\ell)}, Y_{t-1}, J, x, g, \mu, M), \tag{2.12}$$

where $\tilde{\theta}_M^{(\ell)}$ is the ℓ-th draw of $\tilde{\theta}_M$. Then $p(J|Y_t, x, g, \mu, M)$ can be estimated when we substitute (2.12) into Equation (2.8), together with the estimate of $p(J|Y_{t-1}, x, g, \mu, M)$ if $t > \tilde{T}$, or with the uniform prior distribution $p(J|x, g, \mu, M)$ if $t = \tilde{T}$.

2.5. Illustration of posterior distribution of sample size via simulation

In this section, we apply our method to three specific hypothetical scenarios to see what posterior distributions of the sample size we would obtain in each of these scenarios. In the first two scenarios we impose a linear structural break in the data – a break in the middle of the sample and a break near the end of the sample, respectively. Structural breaks in the middle and at the end of the sample are relevant in practice. For example, van Berkum et al. (2013) test multiple structural breaks in the Dutch male data for the years [1950 : 2008] using the Lee-Carter model, and detect breakpoints at year 1970 and 2002. Posterior distributions of sample size with clear peaks are found in these two scenarios. Besides linear structural breaks, sometimes there might exists accelerating trends in historical mortality data, which make the detection of linear structural breaks difficult. To incorporate this situation, we consider an accelerating trend in mortality rates in the third scenario. Posterior distributions of sample size with clear peaks are also found.

The basic idea is as follows. We simulate hypothetical true data, using a certain model specification. We then apply the Bayesian learning approach to determine the posterior sample size distribution. We allow for cases where in the Bayesian learning approach, the user uses a model specification that is different from the model specification that was used to simulate the data. For expositional purposes, we present results only for the case where the data is simulated based on the CBD model specification, whereas the Bayesian learning approach assumes the Lee-Carter model.[13] When simulating the

[13]We have also considered cases where the true data is generated with Lee-Carter, and the Bayesian learning procedure assumes CBD, as well as cases in which there is no model misspecification. In each

hypothetical true data, we impose trend breaks and an accelerating trend via a parabola structure by adjusting the specification of the $\kappa^{(1)}$ process in the CBD model.

We study what posterior distribution of sample size our method would suggest for the Lee-Carter model in each of these situations. We index the simulated data by the years $[T_0 : T] = [1 : 70]$, and use a period of 20 years to update the posterior distributions, i.e., $\left[\widetilde{T}, T\right] = [51 : 70]$. Also, we choose $\mathcal{J} = \{10, 11, ..., 50\}$.

2.5.1. Early linear trend break

Figure 2.4a shows the simulated one year death probabilities based on the CBD model for Dutch males aged 51 to 60 with a linear structural change at year 30. The starting year of the simulated data is 2009. In particular, we impose the following structure on $\kappa^{(1)}_{t,m}$:

$$\begin{aligned} \kappa^{(1)}_{m,t+1} &= \kappa^{(1)}_{m,t} + d_1 + \omega_t, \text{ if } t \leq 30, \text{ and} \\ \kappa^{(1)}_{m,t+1} &= \kappa^{(1)}_{m,t} + 4 \times d_1 + \omega_t, \text{ if } t > 30, \end{aligned} \tag{2.13}$$

with $\omega_t \overset{i.i.d}{\sim} N(0, \sigma^2_\omega)$. The parameter values for d_1 and σ_ω are obtained by fitting the original CBD model to the Dutch male data aged 0 to 90 and years 1970 to 2009. This yields $\hat{d}_1 = -0.0204$ and $\hat{\sigma}^2_\omega = 0.00063$. We update the posterior distribution using simulated data from $[51 : 70]$. The results are shown in the first row in Figure 2.5. The horizontal axis denotes the starting year. The sample size corresponding to a certain starting year and forecast horizon μ is computed in the same way as in Figure 2.1: $J = 70 - \mu -$ *starting year* $+1$. For example, for $\mu = 10$, the starting year 30 is equivalent to a sample size of 31 years (30 to 60). We see that the posterior distributions become flatter and flatter, and the peak of the posterior distribution shifts from around 30 to around 40 as μ decreases from 20 to 1. The reason is that we update the posterior distribution based on the likelihood with respect to data μ years ahead. Therefore, for a smaller μ it is more difficult for the model to distinguish between random variations and the underlying trend in the data, and thus the posterior distributions are less informative. Moreover, data before year 30 receives only small probability mass for all μ-s, meaning that data before the break is considered not informative and is therefore discarded. In other words, the existence of the structural break is detected by our method.

of these cases, the results are similar to the ones presented in this section.

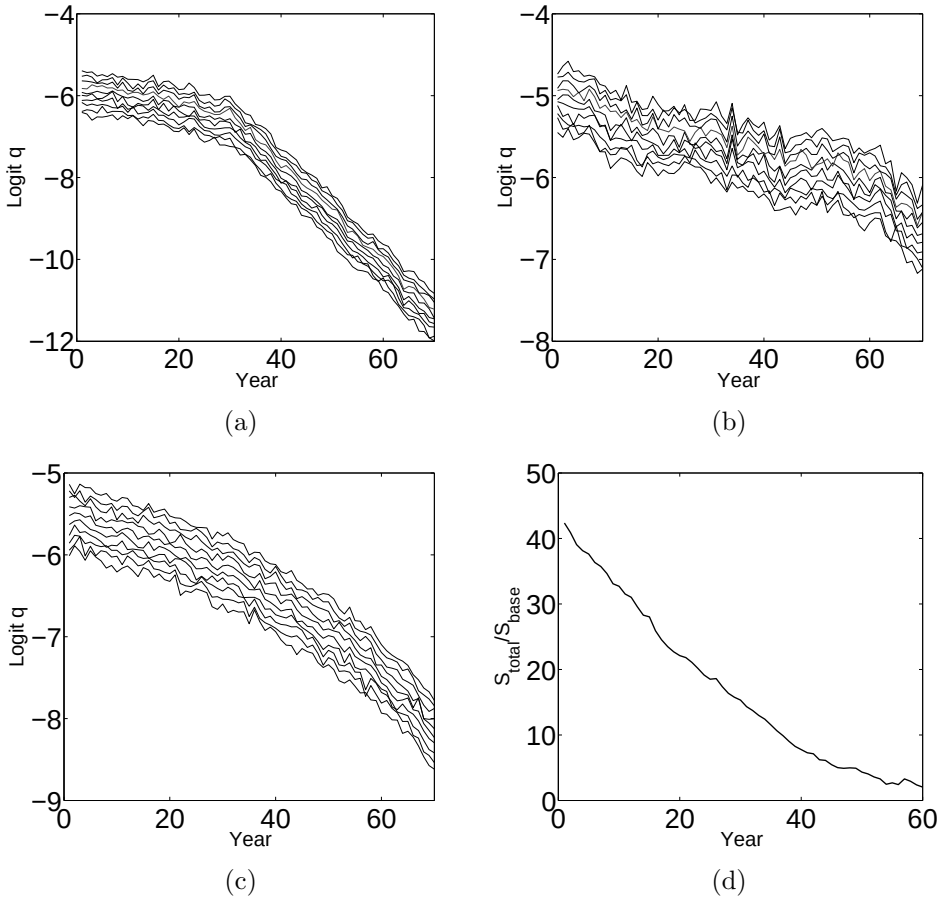


Figure 2.4: Panel 2.4a and 2.4b display the simulated data based on the CBD model with a linear structural break at year 30 and 60, respectively. Panel 2.4c display the simulated data with an accelerating trend. The parameters of the CBD model are estimated using age 0 to 90 and years 1970 to 2009. The starting values of the simulated data are Dutch males aged 51 to 60 in 2009. Panel 2.4d reports the results of the method from Booth et al. (2002) to the non-linear simulated data.

2.5.2. Late linear trend break

Figure 2.4b shows the simulated one year death probabilities based on the model specification as in (2.13), but now with the break date at 60. Again, we update the posterior distribution using the sample $[51 : 70]$, which means that we have only 10 observations after the trend break.

The results are shown in the second row in Figure 2.5. Since we have only 10 observations after the break, the forecast horizons 15 and 20 years are not informative. The interpretation of the results is as follows. For $\mu = 10$, the end of sample in the last update is year 60. In other words, no data after the break is used. In this case, any choice of sample size yields a similar future mortality trend. However, models based on larger sample sizes yield a better in-sample fit, and thus receive a larger probability mass. For $\mu = 5$ and $\mu = 1$, the end-of-sample in the final update is after the break, namely 65 and 69, respectively. Therefore, the model based on a shorter sample size yields more precise forecasts. From the results, we see that shorter sample sizes receive more and more probability mass as we change from $\mu = 10$ to $\mu = 1$, meaning that our method detects the existence of the structural break. However, sample sizes of 40 to 50 years still receive most probability mass, because very short sample sizes may yield a poor fit.

The result in this case study indicates how our method treats a recent structural break. When a structural break is detected, our method would not immediately suggest to use only data after the break. Instead, the posterior distribution shifts gradually in favor of the break as more and more supportive data is observed. When there are enough observations after the break, as in Section 5.1, the posterior distributions become more robust to different μ-s, and suggest that only data after the break should be used.

2.5.3. Accelerating mortality trend

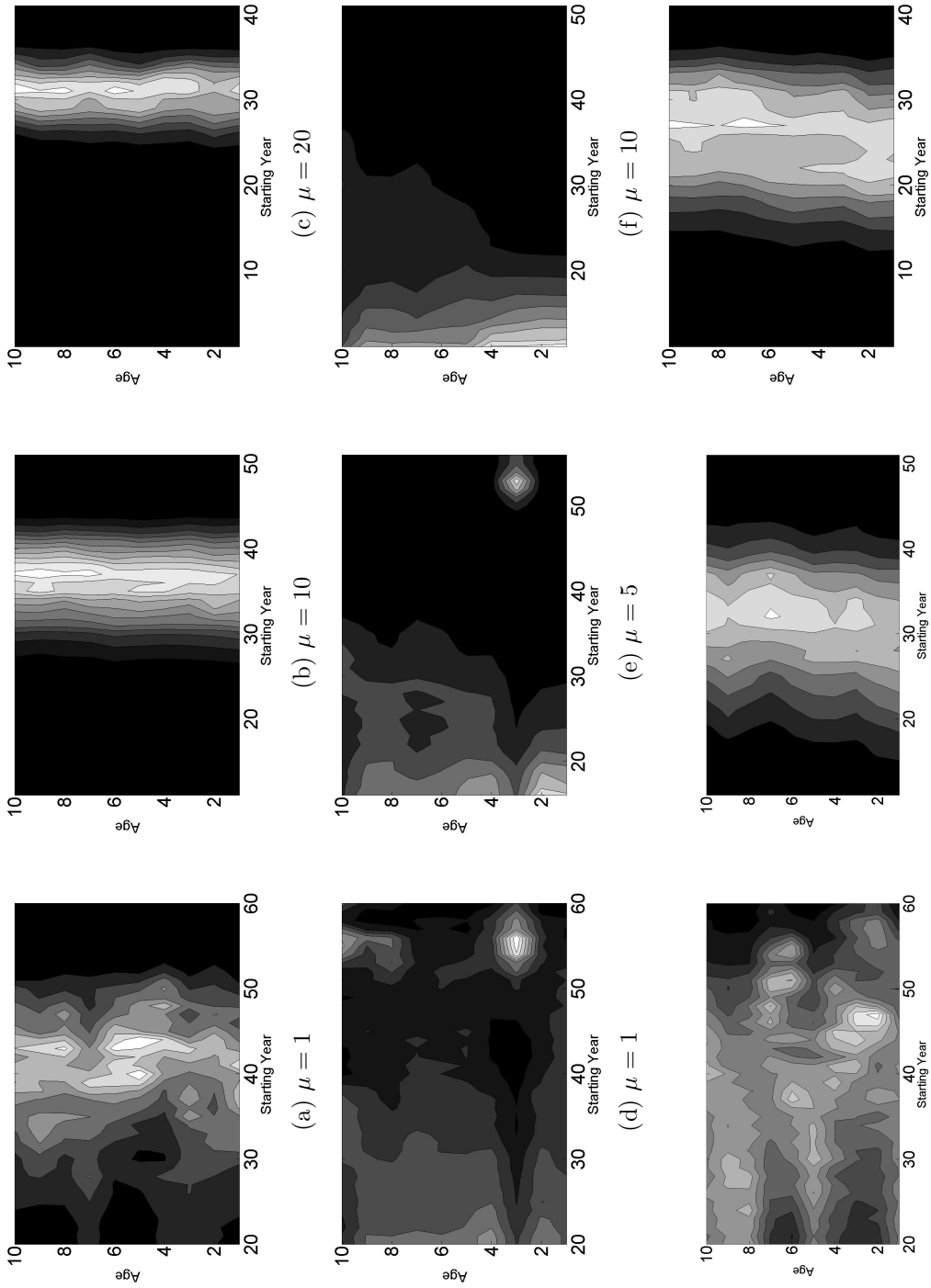
Figure 2.4c shows the simulated data using a CBD model with nonlinear mortality trends. In particular, we impose

$$\kappa^{(1)}_{m,t+1} = a \times \left(\kappa^{(1)}_{m,t}\right)^2 + \kappa^{(1)}_{m,t} + d_1 + \omega_t,\ \omega_t \overset{i.i.d}{\sim} N(0, \sigma^2_\omega), \tag{2.14}$$

with, similarly, $d_1 = -0.0204$ and $\sigma^2_\omega = 0.00063$. Also, we set $a = -0.0075$. The value of a is chosen so that the simulated data has a visual indication of an accelerating trend.

Again, we update the posterior distribution based on the simulated data [51 : 70].

In the case of an accelerating mortality trend, all linear models are misspecified. Figure 2.4d reports the application of the method from Booth et al. (2002) to the data in this case. We see that the ratio S_{total}/S_{base} (as defined in Section 3.2) increases substantially with sample size, thus the Booth's method would suggest the use of the shortest sample size possible in this case. The use of a very short sample size is not desirable, since it may produce estimations and forecasts that are very sensitive to noises in the data. To the contrary, the results in the bottom row in Figure 2.5 show that posterior distributions with clear and concentrated peaks are obtained for all forecast horizons except for $\mu = 1$, and the posterior means of the sample size are of reasonable lengths (around 30 years for $\mu = 10$ and 25 years for $\mu = 20$). The reason is that, in our method, models with smaller sample sizes in general yield more precise point forecasts but poorer fits (larger variances, etc), and our method is able to identify the sample sizes which yield the best balance between the in-sample fit and the out-of-sample forecast performance. However, compared with Section 5.1, we see that the posterior distribution are flatter in this situation, indicating that the accelerating trend indeed leads to more ambiguous choice of sample size.



(a) $\mu = 1$ (b) $\mu = 10$ (c) $\mu = 20$

(d) $\mu = 1$ (e) $\mu = 5$ (f) $\mu = 10$

(g) $\mu = 1$ (h) $\mu = 10$ (i) $\mu = 20$

Figure 2.5: The heat map of the posterior distributions for the Lee-Carter model with $\mu = \{1, 10, 20\}$ for the simulated data with, from top to bottom, a linear trend break at year 30, year 60, and a non-linear trend break based on the CBD model, respectively. Black corresponds to very low probability, light grey corresponds to low probability, and dark grey corresponds to high probability.

2.6. Empirical results

In this section, we derive posterior distributions for the sample size for the mortality rates of the Dutch and the U.S. population. Specifically, we use the one year mortality rates from 1911 to 2009 for the Dutch population and 1933 to 2010 for the U.S. population, i.e., $T_0 = 1911$ for the Dutch case and $T_0 = 1933$ for the U.S. case. For both data sets, we use ages 21 to 89 to fit the Lee-Carter model, and age 60 to 89 to fit the CBD model. Because the Dutch mortality data is available for a longer period, we choose the updating period to be 1980 to 1999, and preserve the last 10 years for the out-of-sample forecast analysis (presented in the next section). However, the U.S. mortality data is not long enough, so we use 1991 to 2010 for the updating, and no out-of-sample forecast is performed. Thus, $[\widetilde{T} : T]$ is $[1980 : 1999]$ for Dutch data and is $[1991 : 2010]$ for U.S data. Moreover, we set $\mathcal{J}_{Dutch} = \{10, 11, ..., 50\}$, $\mathcal{J}_{US} = \{10, 11, ..., 39\}$. For both countries, we determine posterior distributions for the sample size based on forecast horizons $\mu \in \{1, 5, 10, 15, 20\}$. Due to the limitation of space, we report only the results for male data.[14]

2.6.1. Uniform prior

Figure 2.6 reports the conditional posterior distribution of the sample size for Dutch male data. The figure shows that the posterior distributions of the sample sizes can be different for different age and model combinations. Similar to the simulation case studies, we see that in general the posterior distributions become more and more concentrated as μ increases.

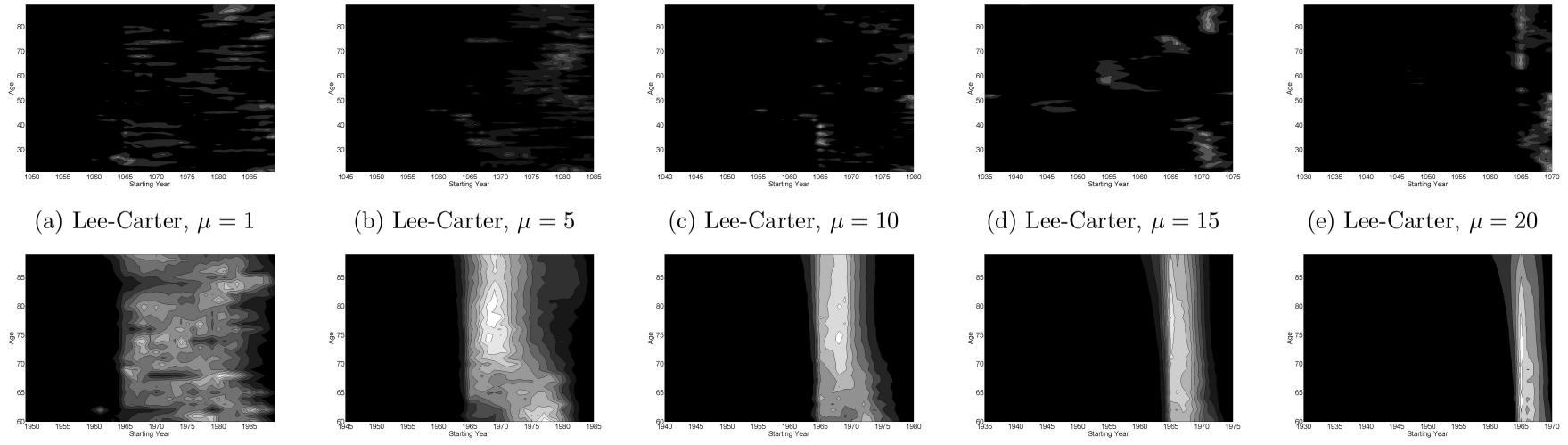
For the CBD model, we see that the posterior distributions are consistent for all μ-s: most probability mass falls within the starting years $[65 : 70]$, especially for $\mu \in \{5, 10, 15, 20\}$. Moreover, the starting years with high probabilities are similar across ages. For the Lee-Carter model, we see that data before the year 1965 receives almost no probability mass, which is in line with the CBD model. However, the posterior distributions for the Lee-Carter model are much more sensitive to the change of μ-s.

Figure 2.7 reports the posterior distributions for the U.S. male population. For the CBD model, the starting years which receive high posterior probabilities are earlier for

[14]The results for female data are available upon request.

the older ages. This result holds for all μ-s. Moreover, the peaks of the posterior distributions for the CBD model move forward as μ decreases, instead of being relatively insensitive to changes in μ as in the Dutch case. One possible explanation is that there exists a cohort effect in the US data. In particular, younger ages have different mortality patterns than older ages, and these patterns are better captured by more recent data. The CBD specification that we use does not model the cohort effect explicitly, and thus treats it as a kind of nonlinearity. As a result, the peak of the posterior distributions might be more sensitive to the changes in μ, which indicates a larger degree of model misspecification. In fact, the existence of a cohort effect in the U.S. data is also confirmed by Cairns et al. (2009).

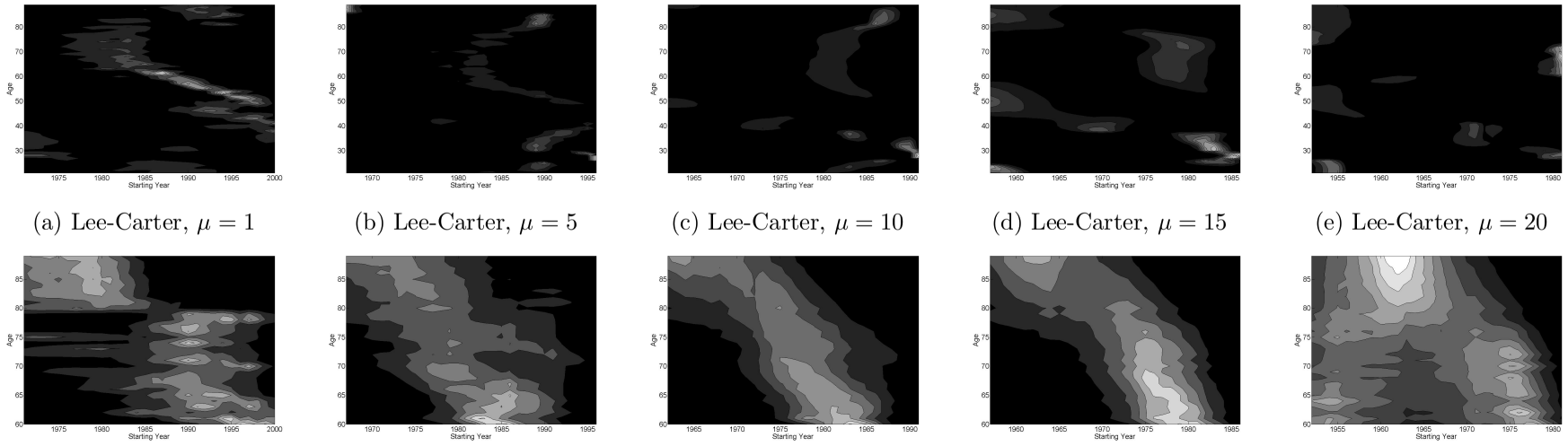
From Figures 2.6 and 2.7, we see that, especially for the Lee-Carter model, the posterior distributions of the sample size are age-specific. This observation is in contrast to the existing sample size determination studies, where typically a single sample size is chosen to estimate the mortality model for all ages. In fact, if such age heterogeneity exists, then one may overlook relevant age-specific information when choosing a single sample size. As an illustration we consider the method by Booth et al. (2002) who determine an optimal sample size from the κ process of the Lee-Carter model. As a sensitivity test, we apply this method to three sub-age groups of the U.S. population. The three groups are $[21 : 40]$, $[41 : 60]$, and $[61 : 90]$. Results are reported in Figure 2.8. We see that while the optimal starting years for the female population are consistent for all three sub-age groups, the optimal starting years for the male population are different. In particular, the optimal starting years for the age group $[41 : 60]$ and $[61 : 90]$ are in mid 1970s and 1960s, indicating that the older ages have an earlier optimal starting year. Moreover, the optimal starting year for the age group $[21 : 40]$ cannot be clearly determined for the male population. Therefore, it seems to make sense to be careful when determining a single starting year for a wide range of ages.



(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 5$ (c) Lee-Carter, $\mu = 10$ (d) Lee-Carter, $\mu = 15$ (e) Lee-Carter, $\mu = 20$

(f) CBD, $\mu = 1$ (g) CBD, $\mu = 5$ (h) CBD, $\mu = 10$ (i) CBD, $\mu = 15$ (j) CBD, $\mu = 20$

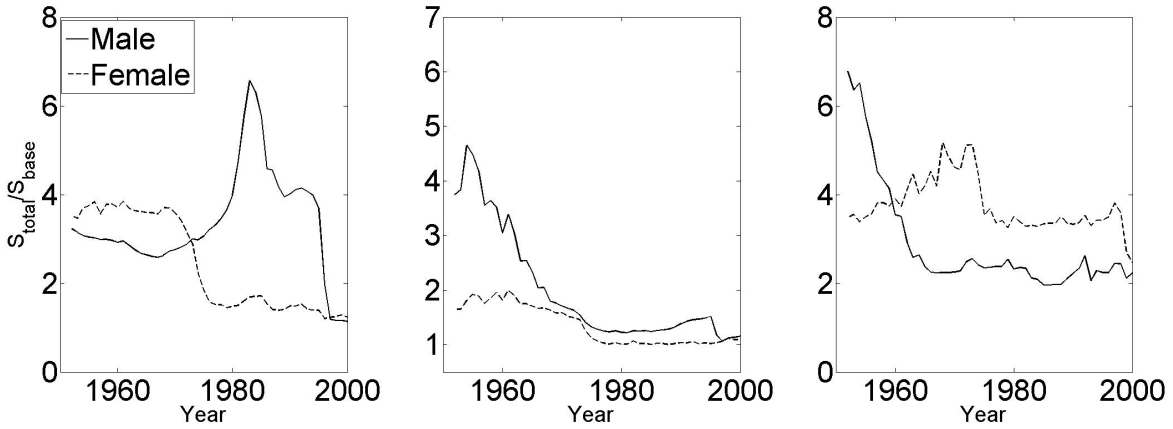
Figure 2.6: The heat map of the posterior distributions of the sample size for $\mu = \{1, 5, 10, 15, 20\}$ for the Dutch male population. The first row is the result for the Lee-Carter model, and the second row for the CBD model. The horizontal axis denotes the starting year. The sample size corresponding to a certain *starting year* and μ is computed as: *sample size* $= 1999 - \mu -$ *starting year* $+1$. Black corresponds to very low probability, light grey corresponds to low probability, and dark grey corresponds to high probability.



(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 5$ (c) Lee-Carter, $\mu = 10$ (d) Lee-Carter, $\mu = 15$ (e) Lee-Carter, $\mu = 20$

(f) CBD, $\mu = 1$ (g) CBD, $\mu = 5$ (h) CBD, $\mu = 10$ (i) CBD, $\mu = 15$ (j) CBD, $\mu = 20$

Figure 2.7: The heat map of the posterior distributions of the sample size for $\mu = \{1, 5, 10, 15, 20\}$ for the US male population. See Figure 2.6 for further details.



(a) U.S. Age [21 : 40] (b) U.S. Age [41 : 60] (c) U.S. Age [61 : 90]

Figure 2.8: The application of the method in Booth et al. (2002) to the sub age groups of the U.S. data. The three sub age groups are [21 : 40], [41 : 60], and [61 : 90], respectively.

2.6.2. Sensitivity with respect to the prior distribution

In this section, we examine the effect of the choice of prior distribution on the corresponding posterior distributions. Due to limitation of space, we report the results of three forecast horizons, 1, 10, and 20 years, for the Lee-Carter model using Dutch data.[15] The results for forecast horizons 5 and 15 years are similar.

For the sensitivity analysis, we consider five sets of prior distributions. Specifically, for each given (x, g, μ, M), we consider $p(J_{i+1}|x, g, \mu, M) = a \times p(J_i|x, g, \mu, M)$, where $J_{i+1} = J_i + 1$ for all i with $J_1 \leq J_i \leq J_{max-1}$, and $a \in \{\frac{1}{1.2}, \frac{1}{1.1}, 1, 1.1, 1.2\}$. In other words, we look at prior distributions for which the prior probability mass increases (decreases) with the sample size. The choices of a allow for great discrepancy among sample sizes. In particular, we have $p(J_{max}|x, g, \mu, M) \approx 50p(J_{min}|x, g, \mu, M)$ when $a = 1.1$, and $p(J_{max}|x, g, \mu, M) \approx 1764p(J_{min}|x, g, \mu, M)$ when $a = 1.2$.

For each model-population-forecast-horizon combination, we

1. obtain the posterior distribution with different prior distributions for each age, then take the average posterior distribution among all ages;

[15] We report the results for the Dutch case since Dutch mortality data appears to be more non-linear than the U.S. data, and is more sensitive to the choice of prior distribution. The results for the Lee-Carter model using U.S. data, and for the CBD model using both data sets are available from the authors upon request.

2. compute the posterior mean, 5% quantile, and 95% quantile for the averaged posterior distributions;

3. compare the sets of posterior mean and quantiles from all prior distributions.

In principle, we can compare the set of posterior mean and quantiles for all single ages. However, the results are qualitatively the same for all ages, and we only report results for the average posterior distribution for the sake of conciseness.

The results are reported in Table 2.1. The first observation is that, for all forecast horizons, as smaller sample sizes receive larger (smaller) prior probability mass, the posterior distribution shifts towards (away from) them. The shift direction of the posterior distribution is consistent with the change of prior distribution. Second, the degree of shift of the posterior distributions is rather small compared to the change of the prior distributions. In particular, as a changes from $\frac{1}{1.2}$ to 1.2 ($p(J_{max}|x, g, \mu, M)$ becomes about 311 thousand times bigger), the posterior mean and quantiles of starting years change only from $(1962, 1969, 1981)$ to $(1977, 1987, 1990)$. This analysis indicates that the sample period is large enough to make the posterior distributions rather robust with respect to the change of prior distributions. Moreover, as the forecast horizon, μ, increases, the posterior distribution becomes more stable. Take the sample example discussed above, but with $\mu = 20$. The posterior distribution changes now from $(1957, 1962, 1967)$ to $(1964, 1968, 1971)$. The degree of change is even much smaller than the change when $\mu = 1$. From the sensitivity analysis, we see that, given the sample period used, the posterior distributions are rather robust to the choice of prior distributions investigated.

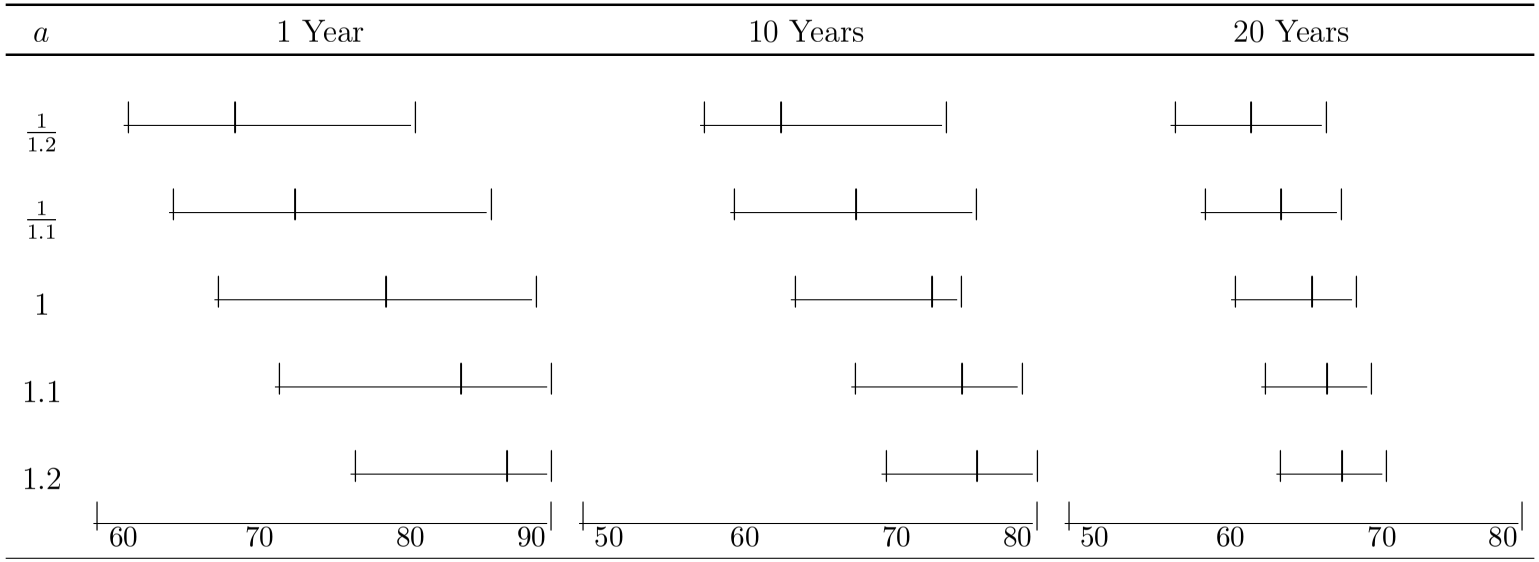


Table 2.1: The posterior mean, 5%, and 95% quantiles of the posterior distributions of sample sizes with different priors. Lee-Carter, Dutch population.

2.7. Out-of-sample forecast

After the posterior distributions $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M)$ and $p(J|Y_t, x, g, \mu, M)$ are obtained for each $\{J, x, g, \mu, M\}$ combination, the corresponding posterior predictive distribution can be derived. For example, for a given (Y_t, x, g, μ, M), we have

$$p(y_{x,t+1,g}|Y_t, x, g, \mu, M) = \sum_{J\in\mathcal{J}} \int_{\tilde{\Theta}_M} p(y_{x,t+1,g}|\tilde{\theta}_M, Y_t, J, x, g, \mu, M) p(\tilde{\theta}_M|Y_t, J, x, g, \mu, M) p(J|Y_t, x, g, \mu, M) d\tilde{\theta}_M. \tag{2.15}$$

Take the Lee-Carter model as an example. Suppose we have $k = 1, 2, ..., K$ draws of $\tilde{\theta}_{LC}$ for each J and μ. Then we have

$$\boldsymbol{\kappa}_{t+1}^{(k)}(J, \mu, LC) \sim \mathcal{N}(\boldsymbol{\kappa}_t^{(k)} + \boldsymbol{d}^{(k)}, \Sigma_\omega^{(k)}), \tag{2.16}$$

and

$$\boldsymbol{y}_{x,t+1}^{(k)}(J, \mu, LC) \sim \mathcal{N}(\boldsymbol{\alpha}_x^{(k)} + \boldsymbol{\beta}_x^{(k)}\boldsymbol{\kappa}_{t+1}^{(k)}, \Sigma_{\varepsilon_x}^{(k)}), \tag{2.17}$$

where all parameters in the above equations are drawn from $p(\tilde{\theta}_M|Y_t, J, x, g, \mu, LC)$, and where the notation (J, μ, LC) emphasizes the dependence of the parameters upon the specific model, sample size, and forecast horizon.[16] The posterior mean of $\boldsymbol{y}_{x,t+1}(J, \mu, LC) = (y_{x,t+1,m}(J, \mu, LC), y_{x,t+1,f}(J, \mu, LC))'$ is thus approximated by

$$\hat{y}_{x,t+1,g}(J, \mu, LC) = \frac{1}{K}\sum_{k=1}^{K} y_{x,t+1,g}^{(k)}(J, \mu, LC), \quad g \in \{m, f\}, \tag{2.18}$$

and the weighted posterior mean, $\hat{y}_{x,t+1,g}(\mu, LC)$, is approximated by

$$\hat{y}_{x,t+1,g}(\mu, LC) = \sum_{J\in\mathcal{J}} p(J|y_t, x, g, \mu, LC) \times \hat{y}_{x,t+1,g}(J, \mu, LC), \quad g \in \{m, f\}. \tag{2.19}$$

where we "integrate" out the sample size. We calculate the posterior predictive distribution similarly for the CBD model.

For the Bayesian model, the mean squared error for period $[T+1 : T+S]$ is calculated

[16] In the sequel, notation such as (μ, LC) indicates an analogous dependence.

as

$$\text{MSE}(x, g, \mu, M) = \frac{1}{S}\sum_{t=1}^{S}(\hat{y}_{x,T+t,g}(\mu, M) - y_{x,T+t,g})^2, \tag{2.20}$$

where $y_{x,T+t,g}$ is the observed mortality data at $T+t$ for (x, g), and $\hat{y}_{x,T+t,g}(\mu, M)$ is the posterior mean of $y_{x,T+t,g}$ according to model M and forecast horizon μ.

We perform the out-of-sample forecast for Dutch data for the period [2000 : 2009], using both the Bayesian models and the original models with the sample size yielding the minimal Deviance Information Criterion (DIC) ratio. The details of the DIC approach are explained in the Appendix.

The sample size which yields the minimal DIC ratio is considered as the optimal one, and is denoted by J_{opt}. In particular, an optimal sample size is computed for each (g, M) combination.[17] The comparison proceeds as follows. For each (g, M) combination, we compute the mean squared error,

$$\text{MSE}(J_{opt}, x, g, M) = \frac{1}{S}\sum_{t=1}^{S}(\hat{y}_{x,T+t,g}(J_{opt}, g, M) - y_{x,T+t,g})^2. \tag{2.21}$$

In Equation (2.21), $\hat{y}_{x,T+i,g}(J_{opt}, g, M)$ is the posterior mean of $y_{x,T+i,g}$ generated by model M and the sample size with the minimal DIC, J_{opt}. The MSE-s generated in (2.21) are then compared with those generated by the Bayesian models.

The J_{opt}-s are reported in Table 2.2. We see that the J_{opt}-s are quite similar for all model-gender combinations. Moreover, they are rather small: They range from 10 years to 14 years. Therefore, it seems that there exist non-linearities in the data, and the fits of the linear Lee-Carter and CBD model deteriorate as the sample size increases. This observation is in line with Figure 2.2a, where the fit of the method in Booth et al. (2002) to Dutch data is plotted.

In Figure 2.9, we report the comparison of the normalized DIC ratios and the average posterior distributions for Dutch males data with $\mu = 1$, 10, and 20 using the Lee-Carter model. The normalized DIC ratios are computed by first taking the negative of the DIC ratios, then normalize them to sum up to 1. In this way, in contrast to the original DIC ratios, the higher the normalized ratio, the better the corresponding sample size. Moreover, the normalized DIC ratios can be also treated as a probability

[17] The forecast horizon does not play a role in the DIC method.

J_{opt}	male	female
LC	1986	1990
CBD	1990	1990

Table 2.2: The optimal starting year under the DIC criterion for Dutch data. In each entry, the first number is the optimal starting year for the males population, and the second number the optimal starting year for the females population.

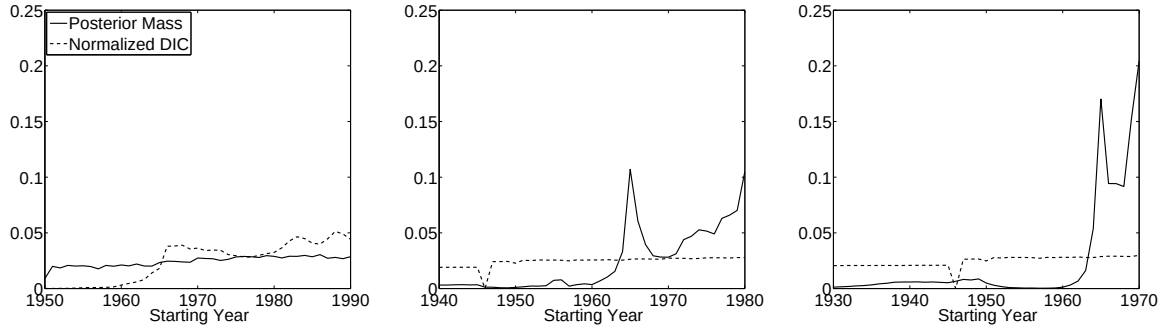
distribution of the sample size, and can thus be directly compared with the posterior distributions. We see that, in all scenarios, the normalized DIC ratios are much flatter than the corresponding posterior distributions, meaning that no particular sample size is clearly better than others.[18] This analysis indicates that, if we would like to determine an optimal sample size based on DIC, the result is likely to be much more sensitive to the possible noise in the data and the chosen sample period compared with the posterior means generated by the Bayesian models. The results for the CBD model are similar and therefore omitted.

Finally, from Figure 2.10, we see that the out-of-sample forecast performance from the models with J_{opt}-s are dominated by the Bayesian model in the majority of cases. For the females population, the Bayesian CBD model performs quite well. In particular, it produces the smallest MSE-s for age 55 and 65, with all μ-s. For the males population, both Bayesian models generate very small MSE-s for the age 85 with all μ-s. The performance of the Bayesian models are between the two models with J_{opt}-s with $\mu \in \{1, 10\}$, and are slightly worse than these two models with $\mu = 20$. However, the posterior distributions of the starting years derived with $\mu = 20$ is more suitable for longer forecast horizons. Therefore, the out-of-sample performance of 10 years does not necessarily mean that the posterior distribution is not suitable to be used for forecasting further into the future.

2.8. Conclusion

In this chapter, we studied the sample size determination of mortality forecasting. In particular, the sample size is incorporated as an extra parameter into the parameter space of the Lee-Carter model and the CBD model, and its conditional posterior distribution

[18] The degree of "flatness" is compared in a heuristic way. We do not formally define a scale of flatness.



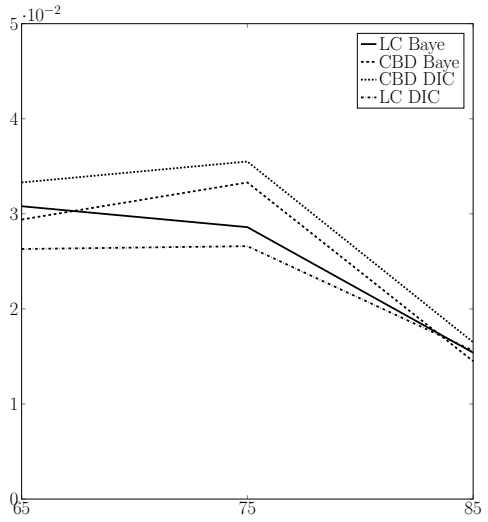
(a) Lee-Carter, $\mu = 1$ (b) Lee-Carter, $\mu = 10$ (c) Lee-Carter, $\mu = 20$

Figure 2.9: The comparison of the normalized DIC ratios (dashed line) and the posterior distributions of sample size (solid line) for $\mu \in \{1, 10, 20\}$ for the Dutch males population using the Lee-Carter model. We first take the negative of the DIC ratios, and then normalized them so that they sum up to 1.

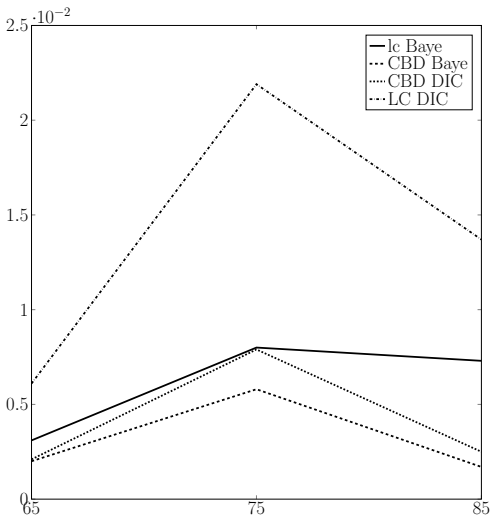
is updated based on historical performance for different forecast horizons. Our method is applicable to many linear mortality models studied in the literature. We focused on the first generation Lee-Carter and CBD model with the underlying latent time effect modeled by a random walk with drift: The simplicity of these two models allows us to illustrate our method more clearly.

We applied our method to Dutch and U.S. data, and found quite concentrated posterior distributions of the sample size for most age, gender, and forecast horizon combinations for both models and data sets. In particular, the posterior distributions of the CBD model are more insensitive to the changes in forecast horizons than the ones from the Lee-Carter model, indicating that the CBD model specification might be more robust to nonlinearities in the data. Moreover, the posterior distributions for both models are age- and gender-specific, meaning that using a single sample size for all ages might not be optimal. In particular, for the CBD model we found that the posterior distribution of the sample size is more concentrated around earlier starting years for older ages. One possible reason is that there exists a cohort effect in the U.S. data: people in younger ages have different mortality patterns than people in older ages, and these patterns are reflected in more recent data. This finding is in line with Cairns et al. (2009).

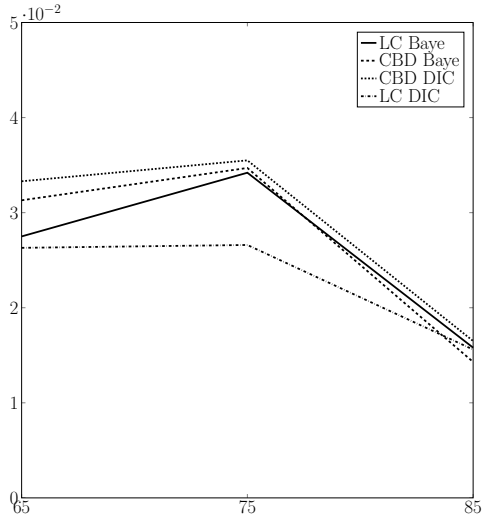
We carried out an out-of-sample forecast analysis using Dutch data from 2000 to 2009. In particular, we compared the out-of-sample performance using the Bayesian models and the original models based on the sample size that gives the minimal Deviance Information Criterion ratio. We found that the forecasts from the Bayesian models are more precise



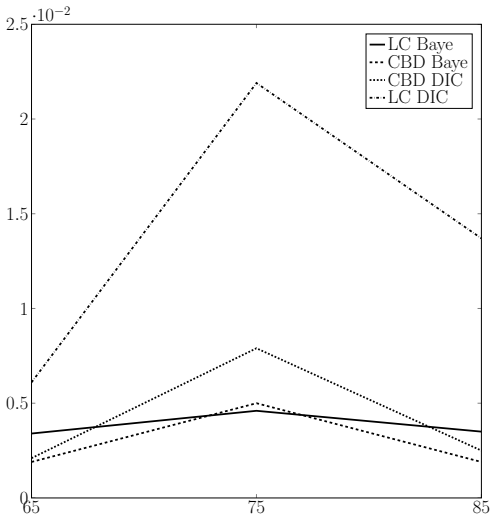
(a) Male, $\mu = 1$



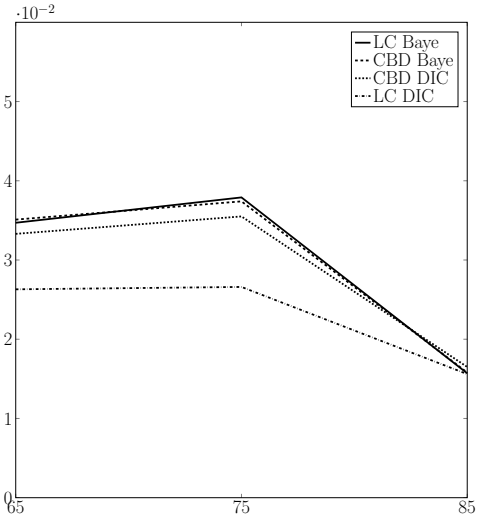
(b) Female, $\mu = 1$



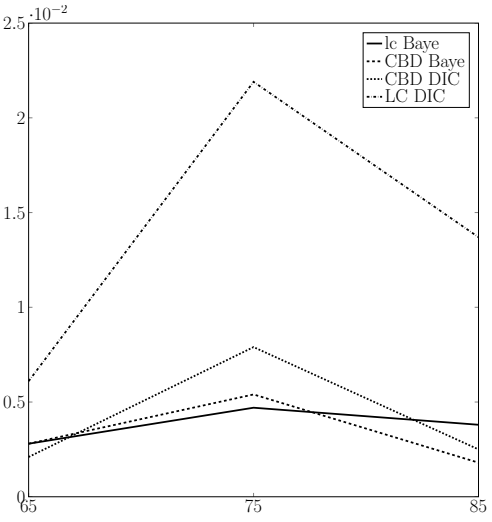
(c) Male, $\mu = 10$



(d) Female, $\mu = 10$



(e) Male, $\mu = 20$



(f) Female, $\mu = 20$

Figure 2.10: The mean squared errors of the out-of-sample forecasts for the Dutch population aged $\{65, 75, 85\}$ for the years 2000 to 2009. The mean squared errors are calculated for the Lee-Carter and the CBD model based on the sample size with the minimal DIC ratios (LC DIC and CBD DIC), as well as the Bayesian models (LC Baye and CBD Baye). All mean squared errors are calculated in terms of logit q.

than the forecasts from the original models in over half of the cases. Therefore, the forecasts of the Bayesian model seem to be more robust than the ones based on the optimal sample size under the traditional Deviance Information Criterion.

There are several directions for future research. First, it could be interesting to apply our method to other mortality models, such as the models studied in Cairns et al. (2011). Second, when more mortality models are included, it could be interesting to compare the performance of each model-sample size combination, for example, using Bayes factors (Kass and Raftery 1995). More insight of the mortality models and the choice of sample size could be gained by doing this analysis. Third, a further model averaging could be done: We can average the forecasts produced by all model-sample size combinations with weights equal to the corresponding likelihood (Hoeting et al. 1999). By doing this, more robust forecasts might be obtained. Finally, the state-space modeling used in this chapter can only deal with linear models with Gaussian innovations. However, it is possible to incorporate particle filters (also known as Sequential Monte Carlo method) into the current framework. When using particle filtering, the mortality model can be nonlinear, and the noise distribution can take any form required. For future research, it would be interesting to explore the optimal choice of sample size in the context of mortality models with Poisson innovations, such the specifications studied in Brillinger (1986), to see which sample sizes give the best combination of in-sample fit and out-of-sample forecasts.

2.9. Appendix

2.9.1. The Gibbs sampler for the mortality models

Lee-Carter model

For each J and μ, $\tilde{\theta}_{LC} = (\boldsymbol{\kappa}', \boldsymbol{d}', \boldsymbol{\alpha}', \boldsymbol{\beta}', \text{vech}(\Sigma_\varepsilon)', \text{vech}(\Sigma_\omega)')'$. In particular, we let Σ_ε be diagonal for all Bayesian models, so $\text{vech}(\Sigma_\varepsilon)'$ is the vector of σ_{ε_x}-s in all cases. Noninformative marginal prior distributions are chosen: $p(\boldsymbol{\alpha}', \boldsymbol{\beta}') \propto 1$, $p(\sigma^2_{\varepsilon_{x,g}}) \propto \frac{1}{\sigma^2_{\varepsilon_{x,g}}}$ for each x and g, and $p(\boldsymbol{d}', \Sigma_\omega) \propto |\Sigma_\omega|^{-\frac{3}{2}}$. Moreover, the marginal prior distributions are assumed to be independent of each other.

Similar to Section 4.2.1, consider the case where we would like to estimate (the poste-

rior distribution of) $\tilde{\theta}_{LC}$ in year $\tilde{T}$. The sample used in the estimation is $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$. The Gibbs sampling proceeds as follows:

1. First we simulate the state process $\boldsymbol{\kappa}$. Denote by D_t the time t information set for $t \in [\tilde{T}-\mu-J+1 : \tilde{T}-\mu]$, $\boldsymbol{\kappa}_t|D_{t-1} \sim N(a_t, R_t)$ the prior of $\boldsymbol{\kappa}_t$ at time t, $\boldsymbol{\kappa}_t|D_t \sim N(\boldsymbol{m}_t, C_t)$ the posterior of $\boldsymbol{\kappa}_t$ at time t, and $\boldsymbol{y}_t|D_{t-1} \sim N(\boldsymbol{f}_t, Q_t)$ the one step forecast distribution.[19] Then the updating equations for time t are

$$\begin{aligned}
\boldsymbol{a}_t =& \boldsymbol{m}_{t-1} + \boldsymbol{d},\\
R_t =& C_{t-1} + \Sigma_\omega,\\
Q_t =& F' R_t F + \Sigma_\varepsilon,\\
A_t =& R_t F Q_t^{-1},\\
\boldsymbol{f}_t =& \boldsymbol{\alpha} + F\boldsymbol{a}_t,\\
\boldsymbol{m}_t =& \boldsymbol{a}_t + A_t(\boldsymbol{y}_t - \boldsymbol{f}_t),\\
C_t =& R_t - A_t Q_t A_t'.
\end{aligned} \tag{2.22}$$

   In system (2.22), $\boldsymbol{y}_t$ is the log central death rate vector at time t, and

$$F = \begin{pmatrix} \boldsymbol{\beta}_m' & 0 \\ 0 & \boldsymbol{\beta}_g' \end{pmatrix},$$

   where $\boldsymbol{\beta}_m$ and $\boldsymbol{\beta}_g$ are column vectors representing the $\boldsymbol{\beta}$-s for male and female, respectively.

2. Fit the updating system (2.22) for each $t \in [\tilde{T}-\mu-J+1 : \tilde{T}-\mu]$, then draw $\boldsymbol{\kappa}_{\tilde{T}-\mu}$ from $\mathcal{N}(\boldsymbol{a}_{\tilde{T}-\mu}, R_{\tilde{T}-\mu})$. For the rest of the t-s, draw $\boldsymbol{\kappa}_t$ from

$$\boldsymbol{\kappa}_t|\boldsymbol{\kappa}_{t+1}, y \sim \mathcal{N}(\boldsymbol{h}_t, H_t),$$

   where $\boldsymbol{h}_t = \boldsymbol{a}_t + B_t(\boldsymbol{\kappa}_{t+1} - \boldsymbol{a}_{t+1})$, and $H_t = R_t - B_t R_{t+1} B_t'$, and $B_t = C_t R_{t+1}^{-1}$.

[19] At the first step, i.e., when $t = \tilde{T}-\mu-J+1$, the prior distribution of $\boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}$ is $N(a_{\tilde{T}-\mu-J+1}, R_{\tilde{T}-\mu-J+1})$, where $a_{\tilde{T}-\mu-J+1}$ and $R_{\tilde{T}-\mu-J+1}$ are the best estimates of the mean and variance of $\boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}$, respectively.

3. For every age x and gender g, draw $\sigma^2_{\varepsilon_x}$ from

$$\sigma^2_{\varepsilon_x}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\kappa},y \sim \text{Inv-Gamma}(\frac{J}{2},\frac{\sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu}(y_{x,t,g}-\alpha_{x,g}-\beta_{x,g}\kappa_{t,g})^2}{2}).$$

4. For each gender g, let $X=(\iota,\boldsymbol{\kappa}_g)$, where ι is a $J\times 1$ vector of ones, and $\boldsymbol{\kappa}_g$ is the vector of $\kappa_{t,g}$-s. Then for each age x, draw $\alpha_{x,g}$ and $\beta_{x,g}$ from

$$(\alpha_{x,g},\beta_{x,g})|y,\boldsymbol{\kappa}_g,\sigma^2_{\varepsilon_{x,g}} \sim \mathcal{N}((X'X)^{-1}X'y,\sigma^2_{\varepsilon_{x,g}}(X'X)^{-1}).$$

5. Draw d from

$$\boldsymbol{d}|\boldsymbol{\kappa},\Sigma_\omega \sim \mathcal{N}(\frac{\boldsymbol{\kappa}_{\tilde{T}-\mu}-\boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}}{J},\frac{\Sigma_\omega}{J}).$$

6. Draw Σ_ω from

$$\Sigma_\omega^{-1}|\boldsymbol{\kappa},\boldsymbol{d} \sim \text{Wishart}(J-1,\frac{\hat{\Sigma}_\omega^{-1}}{J}),$$

where $\hat{\Sigma}_\omega=\frac{1}{J}\sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu}(\boldsymbol{\kappa}_t-\boldsymbol{\kappa}_{t-1}-\boldsymbol{d})(\boldsymbol{\kappa}_t-\boldsymbol{\kappa}_{t-1}-\boldsymbol{d})'$.

We continue the iteration until convergence is reached. Note that in the above sampling procedure, the initial value of m_0 and C_0 should be specified. We let them be $\bar{\boldsymbol{\kappa}}$ and $\hat{\Sigma}_\omega$ obtained from the weight least squared estimation of model (2.2) - (2.3) fitted to $y_{[\tilde{T}-\mu-J+1:\tilde{T}-\mu]}$, respectively.

CBD model

For each J and μ, $\tilde{\theta}_{CBD}=(\boldsymbol{\kappa}',\boldsymbol{d}',\text{vech}(\Sigma_\varepsilon)',\text{vech}(\Sigma_\omega)')'$. Similarly, assume that y_t is a $2n\times 1$ vector for each t. In the context of model (2.10) - (2.11), we have $p=1$, $B_t=0$

and $X_t = 0$ for all t,

$$F_t = \begin{bmatrix} 1 & 1 & ... & 1 & 0 & 0 & ... & 0 \\ x_1 & x_2 & ... & x_n & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 1 & 1 & ... & 1 \\ 0 & 0 & ... & 0 & x_1 & x_2 & ... & x_n \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & ... & 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

for all t, $z_t = [\kappa^{(1)}_{t,m}, \kappa^{(2)}_{t,m}, \kappa^{(1)}_{t,f}, \kappa^{(2)}_{t,f}, d^{(1)}_m, d^{(2)}_m, d^{(1)}_m, d^{(2)}_m,]'$,

$$G_{t,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

for all t, and $\omega_t = [\omega^{(1)}_{t,m}, \omega^{(2)}_{t,m}, \omega^{(1)}_{t,f}, \omega^{(2)}_{t,f}]'$. Also, ε_t is the vector of error terms, Σ_ε is a diagonal matrix, and $\Sigma_{\omega_t} = \Sigma_\omega$ for all t.

Similarly, we choose the noninformative marginal prior distributions $p(\sigma^2_{\varepsilon_{x,g}}) \propto \frac{1}{\sigma^2_{\varepsilon_{x,g}}}$ for each x and g, and $p(\boldsymbol{d}', \Sigma_\omega) \propto |\Sigma_\omega|^{-\frac{3}{2}}$. The marginal prior distributions are assumed to be independent of each other.

Consider the same case as in Section A.1. The Gibbs sampling procedure for the CBD model is

1. For $t \in [\tilde{T} - \mu - J + 1 : \tilde{T} - \mu]$, denote by $\boldsymbol{\kappa}_t | D_{t-1} \sim N(a_t, R_t)$ the prior of $\boldsymbol{\kappa}_t$ at time t, $\boldsymbol{\kappa}_t | D_t \sim N(\boldsymbol{m}_t, C_t)$ the posterior of $\boldsymbol{\kappa}_t$ at time t, and $\boldsymbol{y}_t | D_{t-1} \sim N(\boldsymbol{f}_t, Q_t)$ the one step forecast distribution.[20] The updating equations for time t are

[20]The prior distribution of $\boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}$ is set in the same way as for the Lee-Carter model.

$$\begin{aligned}
\boldsymbol{a}_t =& \boldsymbol{m}_{t-1} + \boldsymbol{d}, \\
R_t =& C_{t-1} + \Sigma_\omega, \\
Q_t =& F' R_t F + \Sigma_\varepsilon, \\
A_t =& R_t F Q_t^{-1}, \\
\boldsymbol{f}_t =& F \boldsymbol{a}_t, \\
\boldsymbol{m}_t =& \boldsymbol{a}_t + A_t(\boldsymbol{y}_t - \boldsymbol{f}_t), \\
C_t =& R_t - A_t Q_t A_t'.
\end{aligned} \tag{2.23}$$

In system (2.23),

$$F = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix},$$

and $X = [\boldsymbol{\iota}, \boldsymbol{x}]'$, where $\boldsymbol{x} = [x_1, x_2, ..., x_n]'$ is the vector of ages. In our case, $\boldsymbol{x} = [61, 62, ..., 89]'$.

2. Fit the updating system (2.23) for all t, and draw the $\boldsymbol{\kappa}$ process in the same way as in the Lee Carter model.

3. For every age x and gender g, draw $\sigma^2_{\varepsilon_x}$ from

$$\sigma^2_{\varepsilon_x} | \boldsymbol{\kappa}, \boldsymbol{y} \sim \text{Inv-Gamma}(\frac{J}{2}, \frac{\sum_t (y_{x,t,g} - \kappa^{(1)}_{t,g} + \kappa^{(2)}_{t,g} x)^2}{2}).$$

4. Draw $\boldsymbol{d}$ from

$$\boldsymbol{d} | \boldsymbol{\kappa}, \Sigma_\omega \sim \mathcal{N}(\frac{\boldsymbol{\kappa}_{\tilde{T}-\mu} - \boldsymbol{\kappa}_{\tilde{T}-\mu-J+1}}{J}, \frac{\Sigma_\omega}{J}).$$

5. Draw σ^2_ω from

$$\Sigma^{-1}_\omega | \boldsymbol{\kappa}, \boldsymbol{d} \sim \text{Wishart}(J-1, \frac{\hat{\Sigma}^{-1}_\omega}{J}),$$

where $\hat{\Sigma}_\omega = \frac{1}{J} \sum_{t=\tilde{T}-\mu-J+1}^{\tilde{T}-\mu} (\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - \boldsymbol{d})(\boldsymbol{\kappa}_t - \boldsymbol{\kappa}_{t-1} - d)'$.

Again, we continue this procedure until convergence is reached. In this chapter, we set the iteration times to be 2,000, and convergence is reached in all cases under the criterion proposed in Gelman and Rubin (1992).

2.9.2. The DIC approach

Define the deviance as

$$D(\tilde{\theta}_M|J,g,M) = -2\log(p(Y_T|\tilde{\theta}_M,J,g,M)) + C \tag{2.24}$$

for each sample size, gender, and model combination, where $\tilde{\theta}_M$ is the parameter drawn from the corresponding posterior distribution, $p(Y_T|\tilde{\theta}_M,J,g,M)$ is the likelihood function, and C is a constant which will cancel out when comparing different models. For the computation of the DIC ratios, we let $J \in \{10,11,...,50\}$ in line with the Bayesian models. Moreover, we use Dutch mortality data from 1950 to 1999 to compute the likelihoods. In other words, $p(Y_T|\tilde{\theta}_M,J,M)$ in (2.24) is the likelihood of $y_{[1999-J+1:1999]}$ based on model M.

Following Spiegelhalter et al. (2002), we define

$$p_D(J,g,M) = E_{\tilde{\theta}_M}[D(\tilde{\theta}_M|J,M)] - D(\hat{\tilde{\theta}}_M|J,M),$$

where $E_{\tilde{\theta}_M}[.]$ is calculated using the posterior distribution of $\tilde{\theta}_M$ given (J,M), and $\bar{\tilde{\theta}}_M$ is the posterior mean of $\tilde{\theta}_M$. The Deviance Information Criterion is then defined as

$$\text{DIC}(J,g,M) = E_{\tilde{\theta}_M}[D(\tilde{\theta}_M|J,g,M)] + p_D(J,g,M). \tag{2.25}$$

Chapter 3

Robust Longevity Risk Management[21]

3.1. Introduction

Pension plans and annuity providers (hereafter referred to as "insurer") are exposed to longevity risk, i.e., the risk due to unanticipated changes in the mortality rates of populations. During recent years, longevity risk has imposed greater and greater financial pressure on the pension and insurance industry due to uncertainty regarding the increase of population life expectancy and economic changes such as lower interest rates and a stricter regulatory environment (Bor and Cowling 2013; Plat 2011). The important role played by longevity risk in asset-liability management calls for more sophisticated risk management methods.

We consider the longevity risk management of an insurer who chooses to hedge the longevity risk using survivor swaps as hedging instruments. Compared to traditional longevity risk transfer approaches, such as buy-outs, hedging using mortality-linked derivatives[22] is cheaper, and may provide more flexibility to the insurer (Cairns et al. 2014). We assume that the insurer uses as objective function the mean-variance or the mean-conditional-value-at-risk of the hedged liabilities. To be able to evaluate the objective function the insurer uses an estimated probability law governing the mortality dynamics. However, the insurer might be concerned about the possible effects of estimation inaccuracy (due to sampling error or model misspecification), when using an estimated probability law. For example, Garlappi et al. (2007) show that mean-variance portfolios of stock indexes may be highly sensitive with respect to parameter estimations, which may lower the portfolios's Sharpe ratios in applications. Similar results are found

[21] This chapter is coauthored with Anja De Waegenaere and Bertrand Melenberg.

[22] Mortality-linked derivatives different from survivor swaps, such as longevity bond or q-forward, can alo easily be incorporated in our model.

by Glasserman and Xu (2013) using commodity futures. Estimation inaccuracy is an important concern when it comes to longevity risk management. In the past decades, many western countries experienced accelerating improvement of life expectancy and/or structural breaks in mortality rates (van Berkum et al. 2013; Li et al. 2013). As a consequence, the estimations and point forecasts produced by mortality models based on linear extrapolation methods (Cairns et al. 2011), probably the most widely used class of mortality models nowadays, may be rather sensitive to the calibration window. For instance, Cairns et al. (2006) obtain substantially different estimation results by fitting the Cairns-Blake-Dowd model to England & Wales males data with two different calibration windows. Therefore, relying only on an estimated mortality model, and hence ignoring the corresponding estimation inaccuracy, might result in a poor performance of one's hedging strategy if the actual probability distribution turns out to deviate from the estimated one.

To be robust against such estimation inaccuracy, we consider optimizing the worst-case value of the objective function, where the worst-case is with respect to a statistical confidence set around the estimated probability law. The confidence set is constructed using the Kullback-Leibler divergence, allowing reformulations of the worst-case optimization problems that can easily be solved, using Ben-Tal et al. (2013) and Ben-Tal et al. (2014). Optimizing the worst-case value of the objective function to deal with estimation inaccuracy is an application of "robust optimization." Robust optimization has been studied extensively in the past two decades, with successful applications to various fields, such as finance, statistics, and engineering (Mulvey et al. 1995; Garlappi et al. 2007; Zhu and Fukushima 2009). For a detailed introduction of the subject, see Ben-Tal et al. (2009).

We apply our robust optimization problem to Dutch male mortality data and compare the performance of the robust optimizations with their nominal counterparts, when the insurer optimizes her portfolio ignoring estimation inaccuracy. We consider the hedging of longevity risk of a few age classes, when survivor swaps for only one age class are available. In the robust optimization we choose our confidence set to deal with sampling error in particular, ignoring possible model misspecification.[23] We find that the robust

[23]Also taking into account model misspecification is straightforward, but would require the use of a much larger confidence set.

optimizations outperform the nominal optimizations in most scenarios when estimation inaccuracies exist, i.e., when the actual probability distribution turns out to deviate from the estimated one.

We also show the impact of population basis risk on the effectiveness of the nominal and robust optimizations. In our analysis we consider hedging using either customized or standardized survivor swaps. A customized swap provides the hedger with payments that match exactly the actual mortality experience in her portfolio; while a standardized swap provides payments that are linked to the mortality experience of a reference population. Although being generally cheaper and having better liquidity, the use of standardized swaps may be less effective than customized swaps due to population basis risk, i.e., the mismatch of the mortality experience in the reference population and the population in the hedger's portfolio. When using standardized instead of customized survivor swaps in our analysis, the presence of population basis risk indeed leads to worse objective function values for both the robust and the nominal optimizations, but the ordering in terms of performance of the nominal and robust optimizations remains unaffected.

This chapter contributes to the rapidly growing literature focusing on longevity risk management, by explicitly taking estimation inaccuracy in the optimization into account. Among the many existing studies,[24] only few deal with estimation inaccuracy. For example, Cairns (2013) derives robust hedging strategies with respect to the re-calibration of the mortality models at a later stage. Cox et al. (2013) derive an optimal portfolio choice given a fixed mortality law, and evaluate the effectiveness of the optimal portfolio choice when the true mortality law is only known to belong to a set of distributions. In contrast, this chapter applies robust optimization to make the longevity risk management robust against estimation inaccuracy.

The remainder of this article is organized as follows. Section 3.2 describes the setup, the construction of the insurer's liabilities, and the survivor swaps. Section 3.3 presents the nominal and robust optimization problems. Section 3.4 reports the application of the robust and nominal optimizations to Dutch male mortality data. Finally we conclude in section 3.5. The appendix contains the reformulations of the robust optimization problems that we used to solve these problems.

[24]For example, Dahl et al. (2008), Li (2014), Cairns (2013), Li and Luo (2012), and Cox et al. (2013).

3.2. Liabilities and swaps

In this section we specify the cash flow of the insurer's liabilities and the considered survivor swaps. We consider two types of survivor swaps: customized swaps and standardized swaps. The floating legs of the customized swap are contingent on the actual mortality experience of the insurer's portfolio specific population, while the floating legs of the standardized swap are contingent on the mortality experience of the reference population. Let $K = \{rp, pp\}$ be the set of populations, where rp denotes the reference population and pp denotes the insurer's portfolio specific population. Denote by $p(t, x, k)$ the probability that a male aged x in year 0 in population k is alive in year t. We shall assume that $p(t, x, k)$ is observed in year t. However, before year t, $p(t, x, k)$ is an unobserved random quantity.

Suppose that at time 0 an insurer has sold n_x units of an annuity to a group of individuals aged x. Each of the annuities involves a commitment to pay 1 euro every year to the annuitant for the rest of his/her life. Denote by X the set of ages of the cohorts to which the annuities are sold, with $|X| = N$. r is the fixed annual risk free interest rate and T is the last year during which a cash flow occurs. T is typically chosen to be, for example, $100 - \min\{x \in X\}$, i.e., a year after which the number of annuitants that survive is negligibly small, so that we can ignore the annuity payments afterwards.

We consider an insurer with a sufficiently large number of annuitants n_x-s, so that the micro-longevity risk (i.e., the risk that the actual survival fractions in population k deviate from the probabilities $p(t, x, k)$) is negligibly small relative to the macro-longevity risk (i.e., the risk that the realizations of $p(t, x, k)$ deviate from their best estimates). Given this assumption, the time 0 (random) discounted liability of the insurer (without hedging) can be written as

$$\tilde{L} = \sum_{x \in X} n_x \sum_{t=1}^{T} \frac{p(t, x, pp)}{(1 + r)^t}. \tag{3.1}$$

At time 0, $\tilde{L}$ is random, since all $p(t, x, k)$ for $t \geq 1$ are random.

As hedging instrument for the longevity risk, we consider survivor swaps (Dowd et al. 2006). Consider a survivor swap contingent on cohort x with maturity T, for $k = rp$ (standardized) or $k = pp$ (customized). Denote by $Fix(x, t, k)$ and $Flt(x, t, k)$

the preset (fixed) and random (floating) payment at year t for $t \in \{1, 2, ..., T-1, T\}$, respectively. At date t the fixed rate payer pays $Fix(x,t,k) - Flt(x,t,k)$ to the counterparty if $Fix(x,t,k) - Flt(x,t,k) > 0$ and vice versa. Following Dowd et al. (2006) we let the fixed and floating payments be given by $Fix(x,t,k) = (1+\tau_x)E_{\tilde{P}}[p(t,x,k)]$ and $Flt(x,t,k) = p(t,x,k)$, respectively, where $\tilde{P}$ is the probability measure used for pricing. In other words, at year t, and given $\tilde{P}$, the preset payment is the time 0 best estimated t-year survival probability of the cohort x multiplied by a number $(1+\tau_x)$, while the random payment is the corresponding realized t-year survival probability. For the formulation and derivation of our optimization problem, we will work with a fixed, constant risk premium.[25] The time 0 (random) discounted cash flows received by the fixed rate payer, $S(x,k)$, can now be written as

$$S(x,k) = \sum_{t=1}^{T} \frac{(1+\tau_x)E_{\tilde{P}}(p(t,x,k)) - p(t,x,k)}{(1+r)^t}. \tag{3.2}$$

In an ideal situation, there are publicly traded survivor swaps for all cohorts $x \in X$. However, at the current stage, derivative products are available only for a few cohorts. To incorporate this fact into our framework, we assume that there is a set $X_S \subset X$ with $|X_S| = m \leq N$ such that tradeable survivor swaps exist only for the cohorts $x \in X_S$.

Aiming to hedge longevity risk, the insurer acts as a fixed rate payer. We denote by a_x the units of swaps for the cohort $x \in X_S$ held by the insurer at time 0, and $\boldsymbol{a}$ the $m \times 1$ vector of a_x-s. We assume that, as a hedger, the insurer does not take short positions in the swap, i.e., $\boldsymbol{a} \in R_+^m$. The insurer's hedged discounted liabilities are given by

$$\widetilde{L} + \sum_{x \in X_S} a_x S(x,k). \tag{3.3}$$

If $X_S = X$, and $k = pp$, the longevity risk can be fully hedged by choosing $a_x = n_x$ for all $x \in X$, and the hedged discounted liabilities becomes $\sum_{x \in X} n_x E_{\tilde{P}}(S(x,pp))$.[26] However, in the general case, when $X_S \neq X$ and/or $k \neq pp$, the longevity risk cannot be fully hedged.

[25] We will examine the impact of the risk premium on the insurer's hedging decision in the numerical study in Section 3.4.

[26] In this chapter we consider only macro longevity risk. However, if micro-longevity risk, i.e., risk related to uncertainty in the time of death if survival probabilities are known with certainty, exists, then the longevity risk cannot be fully hedged even by using customized swaps.

3.3. Optimal longevity risk hedging

In this section we present the insurer's optimization problems, as seen from time 0, when the insurer aims to choose the hedging portfolio $\boldsymbol{a}$ optimally. Let $\boldsymbol{Z}(k)$ be a $(m+1)\times 1$ random vector with the 1-st entry $\tilde{L}$ and the other entries $S(x,k)$, for $x \in X_S$, with the same ordering as in $\boldsymbol{a}$, cf. (3.3). The hedged liabilities can then be written as

$$L(\boldsymbol{Z}(k), \boldsymbol{a}) = (1 \ \boldsymbol{a}')\boldsymbol{Z}(k). \tag{3.4}$$

At time 0 the hedged liabilities are random. As objective function we shall use the mean-variance and the mean-conditional-value-at-risk of the random hedged liabilities. To be able to calculate these objective functions we need the probability distribution $P_{\boldsymbol{Z}(k)}$ of $\boldsymbol{Z}(k)$, which depends on the random survival probabilities $p(t,x,k)$. We assume that the insurer does not know $P_{\boldsymbol{Z}(k)}$, but has to estimate it, using a model like the Lee and Carter (1992)-model. We refer to the estimated distribution as the nominal distribution, denoted by $\widehat{P}_{\boldsymbol{Z}(k)}$. However, the insurer recognizes that her estimates may be subject to estimation inaccuracy (sampling error or model misspecification), and she considers the worst-case objective functions, where the worst-case is with respect to $P_{\boldsymbol{Z}(k)}$ in a confidence set around $\widehat{P}_{\boldsymbol{Z}(k)}$.[27] We present the optimization problems for fixed $k \in \{rp, pp\}$, and write $\boldsymbol{Z} = \boldsymbol{Z}(k)$ for simplicity of notation. Also, we write $P = P_{\boldsymbol{Z}}$ and $\widehat{P} = \widehat{P}_{\boldsymbol{Z}}$.

In the sequel, we shall solve the optimization problems under the assumption that $\boldsymbol{Z}$ has a (multivariate) discrete distribution, represented by the I-dimensional probability vector $\boldsymbol{\pi}$, with components π_i, $i = 1, \cdots, I$, and outcome space $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_I\}$, where $P(\boldsymbol{Z} = \boldsymbol{z}_i) = \pi_i$, $i = 1, 2, ..., I$. We denote by $\hat{\boldsymbol{\pi}}$ the nominal probability vector, so that $\widehat{P}(\boldsymbol{Z} = \boldsymbol{z}_i) = \hat{\pi}_i$, $i = 1, 2, ..., I$. To obtain $\hat{\boldsymbol{\pi}}$, we first fit a (multi-population) mortality model, and use the parameter estimates to simulate the future liabilities and payments of the swaps. Using these simulated values, we then compute $\hat{\boldsymbol{\pi}}$ numerically. In Section 3.4.1 we show how $\hat{\boldsymbol{\pi}}$ can be determined using Lee and Carter (1992) via simulation and

[27] The framework used in this paper is closely related to the multiple preference framework studied by Hansen and Sargent. In Hansen and Sargent (2001), the multiple preference problem characterized by a penalty function is called the *multiplier robust control problem*, while the min-max problem considered in our paper is called the *constraint robust control problem*. As stated in Hansen and Sargent (2001), for each constraint robust control problem, there exists a corresponding multiplier robust control problem, where these two problems have the same solution.

discretization.

Denote by $\hat{\Pi}$ the compact confidence set of probability distributions around $\widehat{P}$. We assume that the insurer now optimizes the worst-case objective function with respect to $P \in \hat{\Pi}$. The optimization problem considered by the insurer is given by

$$\inf_{\boldsymbol{a}\in R_+^m} \max_{P\in\hat{\Pi}} \{E_P(L(\boldsymbol{Z},\boldsymbol{a})) + \lambda \mathcal{R}_P(L(\boldsymbol{Z},\boldsymbol{a}))\} \tag{3.5}$$

where $\mathcal{R}_P$ is a function representing the risk in the liabilities, and λ quantifies the trade-off between the expected value and the risk of the liabilities. The nominal optimization, where the insurer ignores potential estimation inaccuracy and optimizes using $\widehat{P}$, is given by

$$\inf_{\boldsymbol{a}\in R_+^m} \{E_{\widehat{P}}(L(\boldsymbol{Z},\boldsymbol{a})) + \lambda \mathcal{R}_{\widehat{P}}(L(\boldsymbol{Z},\boldsymbol{a}))\}. \tag{3.6}$$

For the risk measure $\mathcal{R}_P$ we consider two cases. In the mean-variance case we take $\mathcal{R}_P = Var_P$ (i.e., the variance of $L(\boldsymbol{Z},\boldsymbol{a})$). In the mean-CVaR case we take $\mathcal{R}_P = \text{CVaR}_{\alpha,P}$, given a confidence level $\alpha \in (0,1)$, where $\text{CVaR}_{\alpha,P}$ is defined by

$$\text{CVaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a})) = \frac{1}{1-\alpha} \int_{L(\boldsymbol{z},\boldsymbol{a}) \geq \text{VaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a}))} L(\boldsymbol{z},\boldsymbol{a}) dP(\boldsymbol{z}), \tag{3.7}$$

with $\text{VaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a}))$, the Value-at-Risk (VaR) of $L(\boldsymbol{Z},\boldsymbol{a})$, defined as

$$\text{VaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a})) = \min\{d \in R | \int_{L(\boldsymbol{z},\boldsymbol{a})\leq d} dP(\boldsymbol{z}) \geq \alpha\}. \tag{3.8}$$

In the literature on robust optimization, (3.5) is referred to as the robust counterpart of (3.6). Robust counterparts are attractive alternatives to the nominal optimization problems, when there is uncertainty about the parameter(s) determining the objective function and/or constraints (Ben-Tal et al. 2009). In our case the parameter is the probability distribution of $\boldsymbol{Z}$.

Before we can solve the robust counterparts, the structure of the uncertainty set, $\hat{\Pi}$, has to be specified. There are many popular structures of the uncertainty set in the literature.[28] We choose our uncertainty set to be characterized by the Kullback-Leibler divergence (also known as relative entropy). This divergence is used in various fields, such as statistics (Reid and Williamson 2011), insurance and financial mathematics (Föllmer

[28]For example, see Zhu and Fukushima (2009) and Ben-Tal et al. (2013).

and Schied 2004; Mania et al. 2005), and macroeconomics (Hansen and Sargent 2001). In our context, the uncertainty set can be written as

$$\hat{\Pi} = \{\boldsymbol{\pi} \in R_+^I | \sum_{i=1}^{I} \pi_i = 1, \sum_{i=1}^{I} \pi_i \log(\frac{\pi_i}{\hat{\pi}_i}) \leq \rho\}, \tag{3.9}$$

for some $\rho > 0$. From (3.9), we see that the degree of divergence between the candidate probability distributions and the nominal one is determined by a single parameter, ρ. Hence, ρ can be interpreted as the degree of ambiguity aversion of the decision maker.

To complete the robust counterpart, we need to choose a specific value for ρ. We consider the case where the distribution P of $\boldsymbol{Z}$ belongs to a parametrized set of probability distributions, such as induced by the Lee and Carter (1992)-model. We choose ρ such that $\hat{\Pi}$ in (3.9) becomes an (approximate) confidence set around $\hat{\boldsymbol{\pi}}$ of at least level $(1-\beta)$. Let $\{P_\theta | \theta \in \Theta \subset R^e\}$ denote this parametrized set of probability distributions, i.e., there exists some $\theta_0 \in \Theta$ such that $P_Z = P_{\theta_0}$. Denote by $\hat{\theta}$ the Maximum Likelihood estimate of θ. Moreover, denote by f_θ the density function of P_θ with respect to some σ-finite measure, μ, and $f_0 = f_{\theta_0}$. The ϕ-divergence between f_θ and f_0 is given by

$$I_\phi(f_\theta, f_0) = \int \phi(\frac{f_\theta}{f_0}) f_0 d\mu.$$

Let $\hat{f}_0 = f_{\hat{\theta}}$. As discussed in Ben-Tal et al. (2013), for the ϕ-s that satisfy certain conditions, the normalized estimated ϕ-divergence

$$\frac{2N}{\phi''(1)} I_\phi(f_0, \hat{f}_0)$$

asymptotically follows a χ^2_e distribution, with the degree of freedom determined by the dimension of the parameter set, Θ. The Kullback-Leibler divergence satisfied the required constraints. Therefore, with ϕ chosen to be the Kullback-Leibler divergence, μ be the Lesbegue measure, and the choice of ρ given by

$$\rho = \frac{\chi^2_{e,1-\beta}}{2N}, \tag{3.10}$$

the set

$$\{\theta \in \Theta | I_\phi(f_\theta, \hat{f}_0) \leq \rho\}$$

is the (approximate) $1-\beta$ confidence set around $\hat{\theta}$. Furthermore, as discussed in Ben-Tal et al. (2013), the uncertainty set of the probability vector, $\hat{\Pi}$, is a "at least" $1-\beta$ confidence set around $\hat{\pi}_i$. In Equation (3.10), $\chi^2_{e,1-\beta}$ is the $1-\beta$ percent critical value of a χ^2 distribution with degrees of freedom e, the dimension of θ, and N is the sample size (used to estimate θ). In this chapter we only consider estimation inaccuracy due to sampling error in $\hat{\theta}$. Also taking into account model misspecification is possible, but it would require a larger choice of ρ, such as $\rho = \frac{\chi^2_{I-1,1-\beta}}{2N}$. For more technical details, we refer to Ben-Tal et al. (2013).

After specifying $\hat{\Pi}$ and ρ, we follow the method proposed in Ben-Tal et al. (2013) and Ben-Tal et al. (2014) to derive reformulations of the robust optimization problems (3.5) which can be solved in a tractable way. These reformulations are presented in the Appendix.

3.4. Performance of the nominal optimization problems and their robust counterparts

In this section we consider an insurer whose portfolio includes annuitants of different age classes and who chooses a hedge portfolio of survivor swaps to hedge the corresponding longevity risk, using one of the optimization problems described in the previous section. We assume that the insurer models the mortality rates in the reference population by the Lee-Carter model (Lee and Carter 1992) and the mortality rates in the insurer's portfolio by the method proposed in Plat (2009).[29] The insurer uses mortality data to estimate the distribution of the mortality law and to construct the corresponding uncertainty set. In this section we evaluate the performance of both the nominal optimization problems and their robust counterparts. We first specify the parametric family of the distribution of the mortality law, the (empirical) nominal distribution, and the uncertainty set determining the robust counterpart. Next, we specify the insurer's portfolio. Finally, we present the performance evaluation of the nominal and robust hedging strategies.

[29] We make these choices for illustrative purposes only. Other models can easily be incorporated in our approach as well. For example, see Cairns et al. (2011) for single population, and Li and Lee (2005), and Dowd et al. (2011) for multi-population mortality modeling.

3.4.1. The parametric family, the nominal distribution, and the uncertainty set

In this subsection we specify the parametric family of the distribution of $\boldsymbol{Z}$. For given $k \in \{pp, rp\}$, $\boldsymbol{Z}$ depends on $p(t, x, k)$ for $t = 1, \cdots, T$ and $x \in X$. We first describe the parametric distribution of $p(t, x, k)$ for $k = rp$ and $k = pp$, respectively. Then we construct the induced parametric family of distributions of $\boldsymbol{Z}$, i.e., the probability vectors π such that $P(\boldsymbol{Z} = \boldsymbol{z}_i) = \pi_i$, $i = 1, 2, ..., I$. Finally, we present the uncertainty set $\hat{\Pi}$, see (3.9).

The reference population—First, we model the mortality process of the reference population, i.e., $p(t, x, rp)$ for $t = 1, \cdots, T$ and $x \in X$. Denote by $m(t, x, rp)$ the one-year crude death rate[30] in year t applying to the cohort whose age is x in year 0 in the reference population rp. The t-year survival probabilities can be approximated by (Pitacco et al. 2009)

$$p(t, x, rp) = \exp(-\sum_{s=1}^{t} m(s, x, rp)).$$

Let $\boldsymbol{y}_t$ bet the N-dimensional column vector with as components $\log(m(t, x, rp))$, $x \in X$. The Lee-Carter model (Lee and Carter 1992) models this vector $\boldsymbol{y}_t$ as[31]

$$\begin{aligned} \boldsymbol{y}_t =& \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(0, \Sigma_\varepsilon) \\ \kappa_t =& d + \kappa_{t-1} + \omega_t, \quad \omega_t \overset{iid}{\sim} N(0, \sigma_\omega^2), \end{aligned} \tag{3.11}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are vectors of parameters, $\boldsymbol{\varepsilon}_t$ and ω_t are mutually independent i.i.d shocks ($\boldsymbol{\varepsilon}_t$ is N-dimensional and ω_t is one-dimensional), assumed to be normally distributed, and Σ_ε and σ_ω^2 are the covariance matrix of $\boldsymbol{\varepsilon}_t$ and the variance of ω_t, respectively. The κ process captures the common time varying trend of the central death rates, and it is modeled by a random walk process with drift term d and volatility σ_ω. Moreover, we assume Σ_ε to be a diagonal matrix.

The portfolio specific population—Next, we model the mortality process of the portfolio specific population, i.e., $p(t, x, pp)$ for $t = 1, \cdots, T$ and $x \in X$. Denote by

[30] Number of deaths over the corresponding exposure.

[31] In general the κ process appearing in (3.11) can be modeled as any $ARIMA(p, d, q)$ process. However, the authors state that a $ARIMA(0, 1, 0)$ serves as a reasonable choice. $ARIMA(0, 1, 0)$ is also the most widely used specification of the Lee-Carter model in the literature of mortality modelling.

$\boldsymbol{f}_t$ the N-dimensional vector containing the ratio of one year death probability of the reference population over the portfolio population at year t. Plat (2009) fits $\boldsymbol{f}_t$ by a one factor linear model

$$\begin{aligned} \boldsymbol{f}_t =& \iota + \boldsymbol{w}\vartheta_t + \boldsymbol{\varepsilon}_t^f \\ \vartheta_t =& \delta + \omega_t^f, \end{aligned} \tag{3.12}$$

where ι is a $N \times 1$ vector of ones, and ϑ_t is modeled as a ARIMA(0,0,0) process with mean δ.[32] Moreover, $\boldsymbol{\varepsilon}_t^f \overset{iid}{\sim} N(0, \Sigma_{\varepsilon^f})$ and $\omega_t^f \overset{iid}{\sim} N(0, \sigma^2_{\omega^f})$. We call (3.12) the mortality factor model.

The nominal distribution—So far, we have specified the parametric form of the distribution of $p(t, x, k)$, for $t \geq 1$, $x \in X$, and for given $k \in \{pp, rp\}$. In the optimization problems we use as nominal distribution the estimated induced distribution of $\boldsymbol{Z} = \boldsymbol{Z}(k)$, for given $k \in \{pp, rp\}$. We obtain this nominal distribution, $\widehat{P}$, with $\widehat{P}(\boldsymbol{Z} = \boldsymbol{z}_i) = \hat{\pi}_i$, $i = 1, \cdots, I$, as follows. First, we estimate the parameter values in (3.11) – (3.12). For the reference population, we use Dutch male mortality data from 1980 to 2009.[33] For the mortality process of the portfolio population, we make use of Plat (2009), who fits model (3.12) to Dutch male mortality data of a collective pension portfolio of Dutch insurers containing about 100,000 male policyholders aged 65 or older (and using the Dutch males population as reference population). The parameter estimates for δ and σ_{ω^f} in (3.12) can be found in Equation (3.2) in Plat (2009). Next, we simulate M realizations of $p(t, x, k)$, for $t \geq 1$, $x \in X$, and for $k \in \{pp, rp\}$, using the two estimated models (3.11) and (3.12). Based on these realizations of $p(t, x, k)$, we construct the corresponding realizations of $\boldsymbol{Z}$.

Finally, we divide the range of the M realizations of $\boldsymbol{Z}$ into I subsets. The values of $\boldsymbol{z}_i$ are obtained by taking the average of all realizations of $\boldsymbol{Z}$ that fall into the corresponding subset i and the values of $\hat{\pi}_i$ are the frequencies of the realizations of $\boldsymbol{Z}$ falling into the corresponding subset i. We use $M = 50,000$ and $I = 1,000$. Since the nominal distribution $\widehat{P}$ is derived from the two estimated models (3.11) and (3.12), we actually have $\widehat{P} = P_{\hat{\theta}}$, with $\hat{\theta}$ the estimated parameters of the two models.

[32] For the detailed definition of $\boldsymbol{f}_t$ and $\boldsymbol{w}$, as well as the choice of $ARIMA$ process for ϑ_t process, we refer to Plat (2009).

[33] The data is downloaded from the Human Mortality Database (http://www.mortality.org/).

The uncertainty set—Our uncertainty set is given by (3.9), with ρ given by (3.10), with e equal to the dimension of θ, the vector of parameters appearing in the two models (3.11) and (3.12). In principle, all these parameters are subject to estimation errors. However, as stated in Cairns (2013), most uncertainty of the (single population) mortality forecasting comes from the estimation of the drift term in the κ process. Also, Lee and Carter (1992) only take the uncertainty from the κ process into account when calculating the confidence intervals of the forecast mortality rates. Therefore, we proceed as if only d and σ_ω in (3.11), and only δ in (3.12) are estimated with possible estimation inaccuracy. Thus, we proceed as if $\theta = (d, \sigma_\omega^2, \delta)'$, where θ is assumed to be estimated by the maximum-likelihood estimator, so that $e = 3$ in (3.10), and $\hat{\Pi}$, given by (3.9), follows given this value of ρ. When there is no population basis risk, the estimation of δ in (3.12) does not play a role, so that we have $\theta = (d, \sigma_\omega^2)'$, with $e = 2$, in this case.

3.4.2. Insurer's portfolio

We assume that the insurer's portfolio consists of five Dutch male cohorts aged 64, 65, 66, 67, and 68 in 2009. The annuities start paying out in 2010, i.e., when the cohorts of the annuitants become 65 to 69, respectively. Only the cohort aged 64 has a corresponding survivor swap available in the market, the payments of which also start from 2010. In other words, $X = \{64, 65, ..., 68\}$ and $X_S = \{64\}$. We write $\tau = \tau_{64}$. We normalize all n_x-s, i.e., the number of annuitant in each cohort, to $n_x = 1$. The maturity of all annuities and of the survivor swap is 30 years. In other words, we assume that no cash flows happen after the oldest cohort reaches age 99. The risk free interest rate is assumed to be constant at $r = 4\%$. We set λ, the risk aversion parameter, to 5 for both the Mean-Variance and CVaR specifications.

As mentioned in Bauer et al. (2010), there is not yet a standard practice of pricing survivor swaps at the current stage. However, the prices of the survivor swaps would clearly affect the optimal hedging strategy. To illustrate this effect, we solve the optimization problems for a set of risk premiums. Dawson et al. (2010) consider the risk premium of a survivor swap contingent on $x = 65$ with maturity 50 years to be around 10% (the annual discount rate is assumed to be flat at 3% and mortality data in England and Wales is used). Since we consider survivor swaps of a shorter maturity and a

higher discount rate,[34] we consider risk premiums $\tau \in \{0, 100, 200, 300, 400, 500\}$ basis points. Finally, we let $\alpha = 95\%$ and $\beta = 5\%$, i.e., we consider the 95% CVaR and a 95% confidence interval of the probability distribution.

3.4.3. Comparison of the nominal and robust optimizations

In this subsection we compare the nominal and robust optimal hedging strategies. First, we compare these hedging strategies, considering both the use of standardized and customized swaps. Next, we illustrate the performance of these hedging strategies, when the actual probability distribution P deviates from the nominal, estimated $\widehat{P}$, which is likely to happen in the presence of estimation inaccuracy.

Figure 3.1 shows the optimal hedging strategies ($\boldsymbol{a}^*$) for the robust (diamonds) and the nominal (asterisks) optimizations. The left and the right panels display the $\boldsymbol{a}^*$-s without and with basis risk, respectively. In all cases the optimal amounts of the swaps decrease as the risk premium increases. In particular, for CVaR the nominal $\boldsymbol{a}^*$ becomes 0 when $\tau \geq 3\%$ and the robust $\boldsymbol{a}^*$ becomes 0 when $\tau \geq 4\%$, both with and without basis risk. Also, in the presence of basis risk, the optimal a^* becomes smaller, holding other factors equal. These results are intuitive, since the swap becomes less attractive as its price increases, and the hedge effectiveness decreases when basis risk is introduced.

Next, we evaluate the performance of the robust and the nominal optimization when the true underlying probability distribution P differs from the nominal distribution $\widehat{P} = P_{\hat{\theta}}$, with $\hat{\theta} = (\widehat{d}, \widehat{\sigma}^2_\omega, \widehat{\delta})$. To do so, we consider a range of different hypothetical "true" probability distributions $P = P_\theta$, allowing for $\theta \neq \hat{\theta}$. For each of these hypothetical true distributions, we evaluate the performance of the robust ($\boldsymbol{a}^*_r$) and the nominal ($\boldsymbol{a}^*_n$) optimal strategies by determining the value of the mean-variance and the mean-CVaR objective functions, i.e., we calculate with respect to the true hypothetical distribution P_θ

$$E_{P_\theta}(L(\boldsymbol{Z}, \boldsymbol{a})) + \lambda \mathcal{R}_{P_\theta}(L(\boldsymbol{Z}, \boldsymbol{a})) \tag{3.13}$$

for $\mathcal{R}_{P_\theta} = \mathrm{Var}_{P_\theta}$ and $\mathrm{CVaR}_{\alpha, P_\theta}$ with $\boldsymbol{a} = \boldsymbol{a}^*_r$ and $\boldsymbol{a} = \boldsymbol{a}^*_n$, respectively.

To generate the hypothetical true distributions for the reference population and for the portfolio-specific population, we let the drift terms in (3.11) and in (3.12) be random

[34] Dowd et al. (2006) show that, as determined by their method, the magnitude of the risk premium decreases as the discount rate increases.

draws from their best-estimate distributions $d \sim N(\hat{d}, \sigma_d^2)$ and $\delta \sim N(\hat{\delta}, \sigma_\delta^2)$. Moreover, we inflate the variance of ω_t in (3.11) by a factor $b \in \{1, 2, 3, 4\}$. For σ_d^2 and σ_δ^2 we use the (estimated asymptotic) variances of $\widehat{d}$ and $\widehat{\delta}$, respectively.[35] The choice of b-values is based on the Bayesian approach from Li et al. (2013) (with the Lee-Carter specification applied to the Dutch males mortality data from 1970 to 2009), who find that the 95% quantile of the posterior distribution of σ_ω^2 is around $3.21\hat{\sigma}_\omega^2$.[36]

For each $b \in \{1, 2, 3, 4\}$, we take 200 random draws of the drift terms d and δ. This yields $200 \times 4 = 800$ values of $\theta = (d, \sigma_\omega^2, \delta)$. For each of these 800 valuess, we determine the value of the mean-variance and the mean-CVaR objective functions given in (3.13) evaluated at $\boldsymbol{a}_n^*$ and $\boldsymbol{a}_r^*$, respectively, using $P = P_\theta$ as true distribution. Because we have six risk premiums, we repeat this procedure for each of these values, so that in total $6 \times 800 = 4800$ comparisons are made for each k. Due to the large number of comparisons, we report our results on a (τ, b, k) basis. For any given (τ, b, k), and for a given specification (mean-variance or mean-CVaR), let V_n and V_r be the vector with the corresponding 200 values of the objective function in (3.13), evaluated at a_n^* and a_r^*, respectively. Tables 3.1a and 3.1b report the mean (standard deviation) of V_r, expressed as percentage deviation of the mean (standard deviation) of V_n, i.e., it reports

$$\frac{f(V_r) - f(V_n)}{f(V_n)} \times 100, \tag{3.14}$$

where $f(\cdot)$ is either the mean or the standard deviation. A negative percentage deviation means a better performance (lower mean or lower standard deviation) for the robust optimization. We see that, both with and without population basis risk, $\boldsymbol{a}_r^*$ produces lower means and standard deviations of the objective function values than $\boldsymbol{a}_n^*$ for all but a few cases when $b = 1$ or $\tau = 0$. Moreover, the relative difference between the robust and nominal mean objective function values increases with b and τ. The percentage deviation is 0 for the CVaR specification when $\tau = 3\%$ and 4% since both $\boldsymbol{a}_r^*$ and $\boldsymbol{a}_n^*$ are 0 in these cases. We test the significance of the nonzero differences for each (b, τ, k)

[35] These are $\widehat{\sigma}_d^2 = \frac{\widehat{\sigma}_\omega^2}{30}$ and $\widehat{\sigma}_\delta^2 = \frac{\widehat{\sigma}_{\omega f}^2}{30}$.

[36] Moreover, Börger et al. (2011) model the mortality development for multiple populations with a stochastic trend model and find that, in order to generate wide enough forecast confidence intervals to include all the extreme historic mortality developments, they have to blow up the volatility of their κ_1 process by 2. This κ_1 process in their stochastic trend model is comparable to the κ process in our Lee-Carter model.

combination, and find that most differences are significant at the 95% level.[37]

The results indicate that the robust optimization yields better optimal objective function values in most situations. The only exceptions are the cases with $b = 1$ or $\tau = 0$, where robust optimization produces overconservative results. When the nominal optimization performs better (when the percentage changes are positive), the degree of outperformance is very small, as the percentage deviations are very close to 0. However, when the robust optimization is better, we see much larger differences in the performance. Moreover, the inclusion of population basis risk increases the optimal values (thus lower the hedge quality) on average by 9.25% for mean-variance and 9.25% for mean-CVaR, indicating a lower hedging quality.[38] However, as shown in Table 3.1a and 3.1b, robust optimizations still perform better than the nominal ones in this case.

3.5. Conclusion

In this chaper we study a robust longevity risk management problem for an insurer with committed annuity payments, who uses survivor swaps as hedging instrument. We consider as objective functions the mean-variance and the mean-conditional-value-at-risk of the hedged liabilities. The insurer recognizes that the best estimated probability law affecting her liabilities may be subject to estimation inaccuracy, and optimizes her portfolio with respect to the worst-case scenario, where the worst case is with respect to the set of possible mortality laws determined by the Kullback-Leibler divergence.

We apply the robust optimization problems to Dutch male data and compare their performance with the corresponding nominal ones, which ignore estimation inaccuracy. We construct various realistic settings, where the estimated mortality distribution deviates from the actual one, and find that the robust optimization yields better results than the nominal optimization in almost all scenarios. Moreover, the degree of the outperformance is higher when the real mortality law is further away from the insurer's estimate. The inclusion of basis risk lowers the insurer's hedge quality by increasing the means and standard deviations of the optimal values of the objective functions, but the robust

[37] The only exceptions are the mean-variance specification with $(b, \tau, k) = (1, 0, 2)$ and both specifications with $(b, \tau, k) = (1, 0, 1)$.

[38] Due to limitation of space, the means and std.-s are not reported here. These numbers are available upon request.

optimization still performs better in this case.

In this study we only consider static robust risk management. It would be interesting to develop dynamic robust risk management strategy following, for example, the robust control framework studied in Hansen and Sargent (2008). Moreover, in this chapter the hedging objectives are framed in terms of hedging the present value of future liabilities. We may consider hedging objectives in term of other quantities, such as the future liability cash flows, or the (weighted) sum of some risk measure of future liabilities.

3.6. Appendix

Mean-CVaR—We shall use the reformulation of the CVaR given by Rockafellar and Uryasev (2002), which is given by[39]

$$\text{CVaR}_{\alpha,P}(L(\boldsymbol{Z},\boldsymbol{a})) = \min_{\xi\in R}\{\xi + \frac{1}{(1-\alpha)}E_P([L(\boldsymbol{Z},\boldsymbol{a})-\xi]^+)\}, \tag{3.15}$$

where $[y]^+ = \max\{y, 0\}$. The corresponding robust optimization problem is given by

$$\begin{aligned}\min_{\boldsymbol{a},\xi}\max_{\boldsymbol{\pi}}\ & \boldsymbol{\pi}'\boldsymbol{L}(\boldsymbol{z},\boldsymbol{a}) + \lambda[\xi + \frac{1}{1-\alpha}\sum_{i=1}^{I}\pi_i u_i] \\ s.t.\ & \boldsymbol{a}\in R_+^m, \xi\in R, \boldsymbol{\pi}\in\hat{\Pi}, \\ & u_i \geq L(\boldsymbol{z}_i,\boldsymbol{a}) - \xi,\ \ u_i\geq 0,\ \ \forall\, i\in\{1,2,...,I\}. \end{aligned} \tag{3.16}$$

The uncertain vector, $\boldsymbol{\pi}$, is linear in (3.16). The derivation of the reformulation of (3.16) that we use is therefore a direct application of Theorem 1 in Ben-Tal et al. (2013). We find as reformulation

$$\min_{\boldsymbol{a},\xi,u,\zeta,\eta}\ \lambda\xi + \rho\zeta + \eta + \zeta\sum_{i=1}^{I}\hat{\pi}_i\exp(\frac{L(\boldsymbol{z}_i,\boldsymbol{a}) + \frac{\lambda}{1-\alpha}u_i - \eta}{\zeta} - 1)$$

$$\boldsymbol{a}\in R_+^m, \xi\in R, \eta\in R, \zeta\geq 0, \tag{3.17}$$

$$u_i \geq L(\boldsymbol{z}_i,\boldsymbol{a}) - \xi,\ \ u_i\geq 0,\ \ \forall i\in\{1,2,...,I\}. \tag{3.18}$$

Mean-variance—The derivation for the robust mean-variance optimization is similar.

[39]The CVaR exists if $E_P|L(\boldsymbol{Z},\boldsymbol{a})| < \infty$. This condition holds if $0\leq\tau_x<\infty$ for all $x\in X_S$.

The tractable reformulation is

$$\begin{aligned} \min_{\boldsymbol{a},\eta,\mathcal{K},\xi} \quad & \eta\rho + \xi + \eta \sum_{i=1}^{I} \hat{\pi}_i \exp\left(\frac{\lambda L^2(\boldsymbol{z}_i, \boldsymbol{a}) + (\mathcal{K}+1)L(\boldsymbol{z}_i, \boldsymbol{a}) - \xi}{\eta} - 1\right) + \frac{\mathcal{K}^2}{4\lambda} \\ s.t. \quad & \boldsymbol{a} \in R_+^m, \mathcal{K} \in R, \xi \in R, \eta > 0. \end{aligned} \tag{3.19}$$

Figures

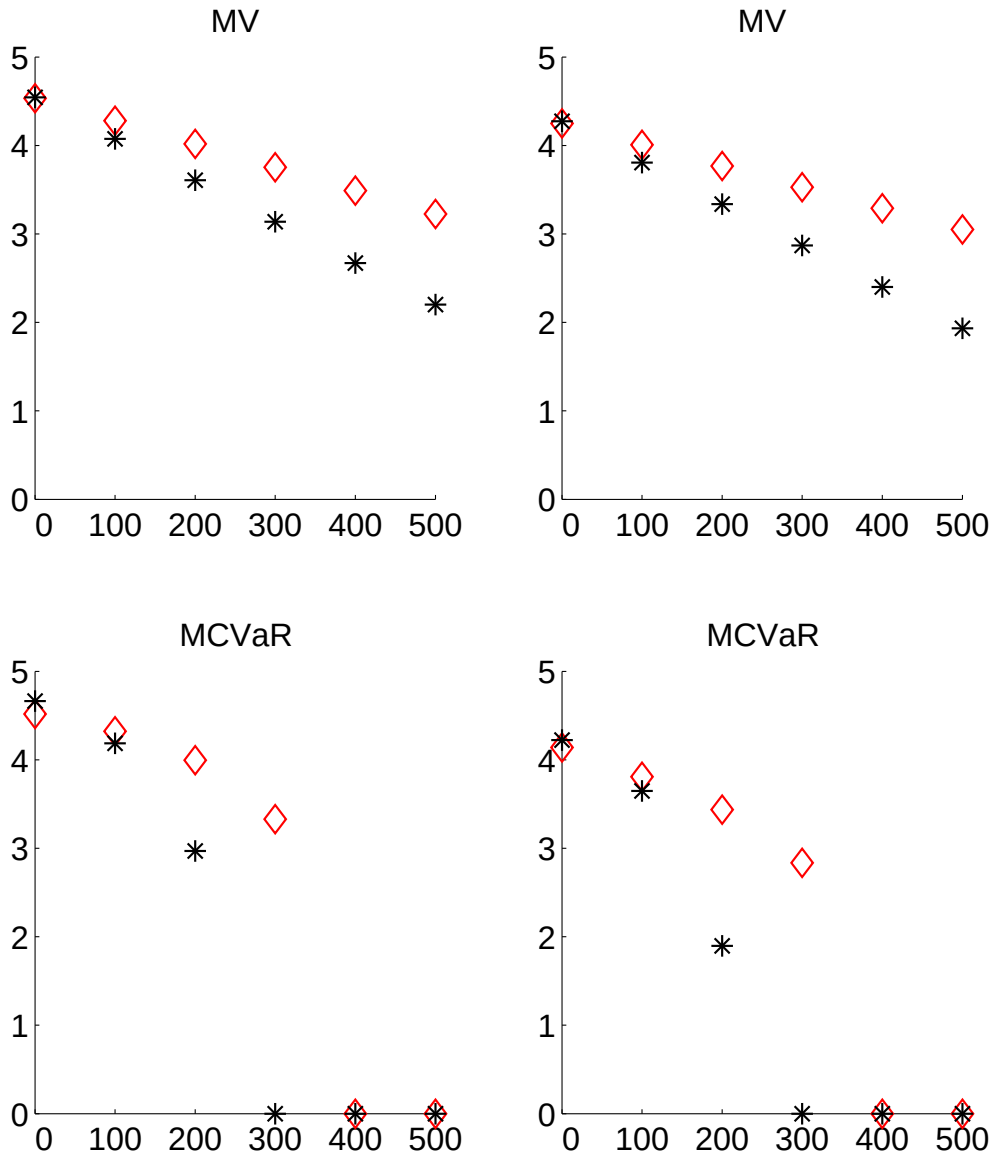


Figure 3.1: The optimal amount of swaps purchased by the insurer as a function of the risk premium in basis points. In each plot the diamonds are the optimal a-s for the robust optimizations (a_r^*) and the asterisks are the optimal a-s for the nominal optimizations (a_n^*). The left panels are without basis risk and the right panel with basis risks. The upper panels show the mean-variance optimal amounts, and the lower ones the mean-CVaR optimal amounts.

	Customized		Standardized		Customized		Standardized	
Risk Premium	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.01	-21.33	-0.01	-0.57	0.01	-24.89	0.0001	-0.98
1%	0.01	-51.39	-0.01	-0.439	-0.01	-51.11	-0.0008	-3.20
2%	0.02	-48.00	-0.05	-11.97	-0.4	-46.18	-0.0036	-7.75
3%	0.06	-46.57	-0.10	-20.06	-0.84	-42.71	-0.0082	-9.60
4%	0.19	-45.29	0.20	-33.40	-0.150	-40.87	-0.0146	-22.07
5%	0.21	-43.89	0.27	-33.72	-0.230	-37.88	-0.0226	-21.33
	$b = 1$				$b = 2$			
0	0.01	-25.24	0.01	-0.37	0.01	-21.90	0.02	-3.04
1%	-0.30	-46.53	-0.26	-2.10	-0.57	-43.77	-0.48	-8.63
2%	-1.12	-42.65	-1.04	-4.26	-2.11	-41.73	-1.97	-4.77
3%	-2.40	-39.11	-2.33	-13.17	-4.43	-39.76	-4.35	-8.78
4%	-4.08	-41.00	-4.05	-18.81	-7.34	-37.71	-7.34	-49.91
5%	-6.06	-39.09	-6.11	-24.99	-10.69	-43.73	-10.81	-67.38
	$b = 3$				$b = 4$			

(a) Mean-variance

	Customized		Standardized		Customized		Standardized	
Risk Premium	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.01	-53.10	0.01	-1.30	0.02	-47.31	0.04	-1.37
1%	-0.03	-5.36	-0.01	-12.63	-0.05	-28.22	-0.08	-10.01
2%	0.02	-36.13	0.12	-41.39	-0.08	-35.69	0.48	-33.92
3%	0.56	-84.40	0.56	-57.39	-0.65	-84.17	0.57	-52.39
4%	0	0	0	0	0	0	0	0
5%	0	0	0	0	0	0	0	0
	$b = 1$				$b = 2$			
0	0.05	-21.33	0.08	1.99	0.07	3.93	0.14	-5.63
1%	-0.82	-27.97	-0.61	-6.04	-0.71	-30.96	-0.42	-7.47
2%	-0.18	-33.54	-0.98	-40.25	-0.27	-32.31	-1.47	-39.18
3%	-2.21	-83.64	-1.66	-59.43	-3.60	-82.55	-2.60	-58.02
4%	0	0	0	0	0	0	0	0
5%	0	0	0	0	0	0	0	0
	$b = 3$				$b = 4$			

(b) Mean-CVaR

Table 3.1: Mean and standard deviation (Std.) of the objective function (3.13) evaluated at a_r^*, expressed as percentage deviation of the mean (in case of the mean) or the standard deviation (in case of the standard deviation) of the objective function (3.13) evaluated a_n^*. The means and standard deviations are calculated over 200 hypothetical true distributions. Panel (a) presents the mean-variance case and panel (b) the mean-CVaR case. The table shows the outcomes for different risk premiums, different b-s, and for $k = pp$ (Customized) and $k = rp$ (Standardized).

Chapter 4

Dynamic hedging of longevity risk: the effect of trading frequency

4.1. Introduction

This chapter considers the question of how to hedge a portfolio of liabilities that are exposed to longevity risk. Longevity risk is the risk due to unanticipated changes in the life expectancy of populations. Longevity risk is becoming a challenge to the pension and annuity industry globally, as pension plans and annuity providers have to make extra payments to the policy holders if their life expectancy increases unexpectedly. In particular, pension plans and annuity providers face substantial risk of making payments longer than anticipated due to the ongoing increase of post-retirement life expectancy. As reported by the Basel Committee's Joint Forum in 2013,[40] if the life expectancy of a typical defined benefit pension fund's members increases by one year, the present value of its liabilities would increase by 3 to 4%. Estimates of the global amount of annuity and pension-related longevity risk exposure range from $15 to $25 trillion (CRO Forum 2010; Blake and Biffs 2012).

A pension plan or an annuity provider (hereafter the hedger) can reduce her longevity risk exposure through the longevity-linked capital market, for example, by trading index-based longevity-linked instruments. An index-based instrument allows the hedger to exchange fixed payments for payments contingent on the actual mortality experience of a reference population, such as the national population.[41] Popular longevity-linked instru-

[40] See the report, Longevity risk transfer markets: market structure, growth drivers and impediments, and potential risks, downloaded from http://www.bis.org/publ/joint34.htm at September 30, 2014.

[41] It is also possible for hedgers to choose customized longevity-linked instruments, by which they exchange fixed payments for payments contingent on the mortality experience in their liabilities, and which can completely eliminate their longevity risk exposure. However, a customized longevity-linked

ments include longevity bonds, q-forwards, and survivor swaps (Menoncin 2008; Dawson et al. 2010). Compared with other types of risk-transfer methods, such as buy-in and buy-out transactions, hedging using longevity-linked instruments can reduce the hedger's cost and provide her with more flexibility (Cairns et al. 2014). In particular, the hedger saves the premium which would otherwise be paid to the insurer. Moreover, the hedger has the opportunity to manage each component of her risk exposure separately, and adjust her hedging strategy over time. However, when hedging using index-based instruments, two things need to be noticed. First, the hedging strategy depends on population basis risk, i.e., the mismatch of the mortality experience of the reference population and the hedger's portfolio-specific population. Second, due to the fact that the longevity-linked capital market is in its early stage of development, the liquidity of longevity-linked instruments is limited at present (Blake and Biffs 2012). As a result, the longevity-linked instruments typically cannot be traded as frequently as other financial assets. Therefore, a realistic hedging strategy using index-based longevity-linked instruments at the current stage should take both the population basis risk and the limited trading frequency into account.

In this chapter we consider dynamic hedging of longevity risk with index-based longevity-linked instruments with limited trading frequency. Due to the limitation of trading frequency, the lack of longevity-linked products, and the population basis risk, the hedger's liabilities cannot be perfectly replicated. Therefore, the longevity risk exposure of the hedger cannot be fully eliminated. Instead, we look at the case where the hedger wishes to minimize her longevity risk exposure. Formally, the hedger wishes to minimize, in a time-consistent way, the variance of her hedging error, which is defined as the deviation of the market value of her investments (in longevity-linked instruments and other financial assets) from the market value of her liabilities, at a specific future valuation date. The variance criterion is commonly used by researchers and practitioners in static settings. However, in the dynamic setting, the variance criterion is also easier to interpret than utility functions. For example, we can measure the hedging quality by looking at the optimal variance of the hedging error, without resorting to the functional form of the utility functions and the choice of risk aversion parameters. The latter is not

instrument would be more expensive and more illiquid (Li and Luo 2012), and such instruments are only available for large risk holders, e.g., the ones with liabilities exceeding around £100 million (Cairns 2013). Therefore, we consider only index-based longevity-link instruments in this chapter.

straightforward to specify for a pension plan. Moreover, the use of the variance criterion is relevant for hedging longevity risk. An insurance company which sell both annuities and life insurance products face longevity risk in two directions. The company loses profits from the life insurance products in case of unexpected mortality deteriorations, e.g., catastrophe events, while it loses profits from the annuities when the life expectancy of their annuitants turn out to be higher than expected. Therefore, the insurance company would like to minimize the deviation of its hedging error from both directions. A similar setup of value hedging is considered in Cairns (2013) and Cairns et al. (2014) in a static framework. A time-consistent strategy, as explained in Strotz (1956), is a strategy that is chosen at the initial date and will be followed by the hedger at any time and in any state of the world, when the hedger re-optimizes at later moments using the same objective function. The variance criterion is widely used in various economic context, however, no time-consistent solutions to these problems in a multi-period setup were proposed until Basak and Chabakauri (2010, 2012). Previous literature only characterizes pre-commitment strategies, which optimizes the criterion at the initial date under the assumption that the hedger will commit herself to follow the initial optimal strategy at later dates. In this chapter, instead of pre-commitment strategies, we follow Basak and Chabakauri (2010, 2012) to derive time-consistent optimal strategies to the minimum-variance criterion, which will be followed when the hedger to re-optimizes the criterion at later dates. The hedger we consider is commonly called "sophisticated" time-consistent in the literature (Grenadier and Wang 2007).

Despite the importance of longevity risk management, most of the existing studies focused on static hedging strategies (Cairns et al. 2014; Li et al. 2014; Li and Luo 2012; Li and Hardy 2011; Blake et al. 2006), which might not take full advantage of the hedging potential of the capital market. The development of dynamic hedging strategies is hindered by the fact that for realistic, discrete-time mortality models, it is difficult to model the evolvement of the value of the longevity-linked contracts over time without resorting to nested simulation (Cairns 2011). Among the few studies which consider dynamic management of longevity risk, Dahl et al. (2011) and Wong et al. (2014) study hedging strategies under continuous-time, multi-population frameworks, while Cairns (2011) considers a discrete-time, single population framework. In particular, Dahl et al. (2011) study a quadratic loss function using a general forward mortality setup, and

consider the limited trading frequency of the longevity-linked instruments. Moreover, Wong et al. (2014) derive time-consistent solutions to a mean-variance problem, but consider only continuous trading. Dahl et al. (2011) and Wong et al. (2014) obtain closed-form solutions of complicated hedging problems. However, there is no empirical support for their mortality models yet, and it is not clear how their hedging strategies would perform in practice.

This chapter contributes to the literature by analyzing the effect of trading frequency in dynamic time-consistent minimum-variance longevity hedging. In particular, we extend the method proposed by Basak and Chabakauri (2010, 2012) to the situation where a part of the assets can only be traded at a limited, deterministic frequency, and obtain semi-closed-form optimal hedging strategies under a forward interest rate and mortality rate framework. We find a direct analog between the optimal strategies in the continuous trading case and the constrained case. Moreover, closed-form solutions can be obtained under reasonable parametric assumptions of the interest rate and mortality rate model. We evaluate the hedging effectiveness in a numerical illustration, where a one-factor Hull-White specification is used for the interest rate process, and a three-factor Hull-While process is used for the mortality process. For the interest rate model, we use parameter estimates from Driessen et al. (2003), who use U.S. interest rate data. For the mortality model, we use the estimates in Blackburn and Sherris (2014), who fit a multi-population mortality model to Australian and Swedish male data. From the numerical study we find that mild trading frequency constraints, such as a two-year frequency, leads to only a slight loss (about 3.7%) of the hedging performance compared to continuous trading. Moreover, even when the longevity bond can only be traded at a very low frequency, such as a five-year frequency, a dynamic hedge still leads to 23% lower hedging error than the static hedge, where the hedger has a constant holding of longevity bonds.

The rest of the chapter is organized as follows. Section 2 gives an illustration of longevity risk and describes the hedger's optimization problem. Section 3 describes the hedger's assets and liabilities. Section 4 gives the optimal hedging strategy to the benchmark minimum-variance optimization. Section 5 gives the optimal hedging strategy to the constrained minimum-variance problem. Section 6 gives a numerical illustration of the hedging strategy and hedge effectiveness. Finally, conclusions are provided in Section 7.

4.2. Longevity Risk and the hedger's problem

Before introducing the hedger's problem, we first illustrate the existence of longevity risk. Figure 4.1 shows the one-year death probabilities, i.e., the probability that an individual dies during a calender year, of the male (left) and female (right) populations aged 65 from 1960 to 2009 in the U.S., the Netherlands, Australia, and Sweden, normalized such that the ones in year 1960 are equal to one.[42] From the figure, we see that there is a downward trend in the one year death probabilities in all gender and country combinations, indicating that the mortality experience is improving in these countries in the past 50 years. More importantly, we see that the development of the death probabilities are volatile for all countries: the death probabilities follow country and age-specific trends, with non-negligible fluctuations in and around the trend. The randomness associated with the future death probabilities is the cause of longevity risk. The ground-breaking paper by Lee and Carter (1992) started a rapid development of stochastic mortality models (Cairns et al. 2011). However, at the current stage, it is still a difficult task to obtain accurate forecasts of future mortality rates. Therefore, nowadays longevity risk is a non-negligible risk component for pension plans and annuity providers.

We consider the case where the hedger can invest in a money market account, a set of zero coupon bonds, and a set of zero coupon longevity bonds contingent on cohort x_0 (hereafter longevity bond).[43] Before introducing the hedger's problem, we first introduce some definitions and notations. Denote by $k \in \{rp, pp\}$ the set of populations. In particular, $k = rp$ refers to the reference population of the longevity bond, and $k = pp$ refers to the population in the hedger's portfolio. A longevity bond is a zero coupon bond with a random principle repayment, which depends on the actual cumulative survival probability of a certain cohort in the reference population, at maturity. The survival probability, the zero coupon bonds, and the longevity bonds are defined as follows.

- ${}_sp(x,t,k)$: the (future) probability that an individual aged x at time t in population k survives up to time $t+s$, given that he is alive at time t. The survival probability, ${}_sp(x,t,k)$, will be observed at time $t+s$, but is random beforehand.

[42] The data is downloaded from the Human Mortality Database: www.mortality.org/.

[43] Other popular longevity-linked derivatives, such as longevity swaps and q-forwards (Dawson et al. 2010), can also be incorporated in our framework. For example, a forward can be mathematically regarded as the exchange of the principle repayment of the longevity bond and a preset payment at maturity.

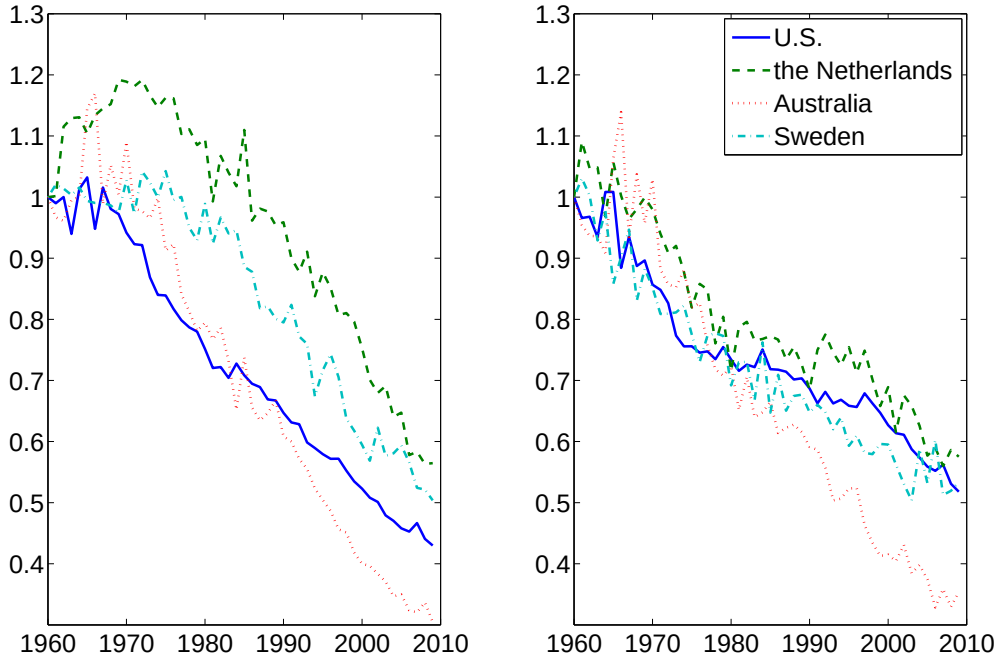


Figure 4.1: One-year death probability of the male (left) and female (right) populations aged 65 for year 1960 to 2009 in the U.S., the Netherlands, Australia, and Sweden. For each country and gender, the death probability in year 1960 is normalized to one.

- $B(t)$: the time t value of the money market account, with $B(0) = 1$.
- $B(t, T)$: the time t price of the zero coupon bond which pays 1 euro at time T with $B(t) = B(t, t)$.
- $L(t, T, x)$: the time t price of the longevity bond contingent on cohort x, which pays ${}_Tp(x, 0, rp)$ euros at time T.

In this chapter, we consider a setting where the hedger implements a dynamic hedging strategy at time 0, and wishes to assess the financial outcome of the hedging strategy at a specific future valuation date, T_0. Without loss of generality, we consider a stylised hedger with N male participants (pension members or annuitants) in a single cohort aged x_0 at year 0.[44] Starting from year T_0, the hedger pays each member one euro per year until the member dies or a terminal date, T_1, is reached. Moreover, we assume that N is large enough, so the number of survivors in year t can be reasonably approximated by $N \times {}_tp(x_0, 0, pp)$. In other words, we focus only on macro longevity risk. Cairns (2013) and Cairns et al. (2014) consider a similar setting, but study static rather than dynamic

[44] Our setup can be naturally generalized to multiple cohorts and participants of both genders.

hedging strategies. Similar to Cairns (2013) and Cairns et al. (2014), we assume that the hedger wishes to minimize the variance of her hedging error, defined as the deviation of the market value of her investments (in longevity-linked instruments and other financial assets) from the market value of her liabilities, at T_0. We assume that there are M zero coupon bonds with different maturities, collected in the set $\mathcal{T} = \{\tau_1, \tau_2, ..., \tau_M\}$, available at any time t. However, due to the fact that the longevity-linked capital market is recently developed, and there are not yet many tradable products, we assume that there exist only $\tilde{M} \leq M$ longevity bonds available in the market, with $\tilde{\mathcal{T}} \subset \mathcal{T}$.

In order to define the dynamics of mortality rates, interest rates, and other relevant quantities, some preliminary setups are necessary. Similar to, for example, Wong et al. (2014), we fixed a finite horizon T^*,[45] and a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, where $\mathcal{F}_0$ is augmented by all the P-null subsets of $\mathcal{F}$, and $\mathcal{F}_t$ is the σ-field generated by an $N = n_r + n_\mu$-dimensional standard Brownian motion under P, $W^P(t) = (W_r^P(t)', W_\mu^P(t)')'$. All stochastic processes are assumed to be well-defined, and adapted to $\{\mathcal{F}_t, t \in [0, T^*]\}$. The evolvement of the zero coupon bonds is driven by the n_r-dimensional process $W_r^P(t)$, while the evolvement of the observed mortality rates is driven by the n_μ-dimensional process $W_\mu^P(t)$. Furthermore, we assume independence of $W_r^P(t)$ and $W_\mu^P(t)$ for all t, i.e., the evolvement of mortality rates is assumed to be independent of the financial markets. We model the money market account as

$$dB(t) = r(t)B(t)dt, \quad B(0) = 1, \tag{4.1}$$

where $r(t)$ is the instantaneous spot rate at time t. We assume the existence of a risk neutral measure, Q^λ, with the money market account as the numéraire, by which prices are determined. The existence of Q^λ guarantees the exclusion of arbitrage opportunities. In particular, Q^λ is characterized by the Radon-Nikodym density

$$(\frac{dQ^\lambda}{dP})_t = \exp(-\int_0^t \lambda(s)' dW(t) - \frac{1}{2}\int_0^t \|\lambda(s)\|^2 ds), \tag{4.2}$$

where $\lambda(t) = (\lambda_r(t)', \lambda_\mu(t)')'$ is the market price of risk vector with respect to Q^λ, satisfying appropriate regularity conditions (See, for example, Proposition 1.7.31 in Jeanblanc

[45] We assume that $T^* \geq \max\{T_1, \tau_M\}$, where τ_M is the longest maturity of the zero coupon bond (and thus the longevity bond).

et al. (2009).). In this chapter, we assume that the zero coupon bond market is complete, so $M = n_r$. However, we allow for the situation where $\tilde{M} < n_\mu$, in which case the longevity bond market is incomplete.[46] For simplicity of notation, we write Q^λ as Q in the sequel. Let $W^Q(t) = (W_r^Q(t)', W_\mu^Q(t)')'$ be an N dimensional process satisfying

$$dW^Q(t) = dW^P(t) + \lambda(t)dt. \tag{4.3}$$

By Girsanov's Theorem (Karatzas 1991), $W^Q(t)$ is an N-dimensional Brownian motion under Q. We assume that $\lambda_r(t)$ only depends on information regarding the financial markets, and $\lambda_\mu(t)$ only depends on information regards the mortality processes. As a result, $W_r^Q(t)$ and $W_\mu^Q(t)$ are independent for all t.

For simplicity of notation, define $Y(t)$ as the normalized time t discounted market value of the hedger's liabilities, i.e., the hedger's time t market value of liabilities is $N \times Y(t)$ euros. For $t \in [0, T_0]$, $Y(t)$ can be written as

$$\begin{aligned} Y(t) =& {}_tp(x_0, 0, pp) E_t^Q[B(t, T_0) \int_{T_0}^{T_1} B(T_0, s)_{s-t}p(x_0 + t, t, pp)ds] \\ =& E_t^Q[\int_{T_0}^{T_1} B(t, s)_s p(x_0, 0, pp)ds]. \end{aligned} \tag{4.4}$$

The hedger's initial assets are assumed to be equal to her initial market value of liabilities, i.e., $w(0) = Y(0)$.[47] At time t, her wealth is

$$w(t) = u_0(t)B(t) + \boldsymbol{u}_1(t)'\boldsymbol{L}(t) + \boldsymbol{u}_2(t)'\boldsymbol{B}(t) \tag{4.5}$$

with the self-financing budget constraint

$$dw(t) = u_0(t)dB(t) + \boldsymbol{u}_1(t)'d\boldsymbol{L}(t) + \boldsymbol{u}_2(t)'d\boldsymbol{B}(t). \tag{4.6}$$

[46] The case where the zero coupon bond market is incomplete can be incorporated naturally. When the longevity bond market is incomplete, λ cannot be uniquely determined by market prices. In this case, we assume that λ is determined by the market clearing conditions, together with some underlying no-arbitrage equilibrium conditions. For methods of determining an equivalent martingale measure, see, e.g., Kallsen (2002).

[47] We allow initial assets that are not equal to $Y(0)$. In fact, in this paper we do not impose any solvency constraint on the hedger. Therefore, the hedger's wealth does not need to be above some threshold. Moreover, as will be seen later, the optimal variance-minimizing hedging strategy in both the benchmark case and the constrained case does not depend on $w(t)$ for all t. The only requirement for the optimization to make sense is that the wealth process is integrable throughout the planning horizon.

In (4.5) - (4.6), $u_0(t)$, $\boldsymbol{u}_1(t)$, and $\boldsymbol{u}_2(t)$ are the hedger's holdings of money market account, longevity bond, and zero coupon bond at time t, respectively. Moreover, $\boldsymbol{L}(t)$ and $\boldsymbol{B}(t)$ are the $\tilde{M}$ and M dimensional vector containing the time t price of the longevity bonds and the zero coupon bonds with maturities in $\tilde{\mathcal{T}}$ and $\mathcal{T}$, respectively. Given the above definition, the time T_0 hedging error of the hedger is $w(T_0) - Y(T_0)$, and the objective of the hedger can be formulated as

$$\min_{\boldsymbol{u}\in\mathcal{U}} \ \mathrm{Var}_0[e^{-\int_0^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))], \tag{4.7}$$

subject to the budget constraint (4.6), where $\boldsymbol{u}(t) = (\boldsymbol{u}_1(t)', \boldsymbol{u}_2(t)')'$. $\mathcal{U}$ is the set of admissible strategies, i.e., each $\boldsymbol{u} \in \mathcal{U}$ is $\mathcal{F}$-predictable, and satisfies standard integrability conditions. We consider the variance of the discounted hedging error since, as mentioned in Basak and Chabakauri (2010), optimizing with the discounted value would significantly facilitate the derivation of the dynamic hedging strategies. We consider the hedging problem in two cases: In the benchmark case, both the zero coupon bonds and the longevity bonds can be traded continuously; while in the liquidity constrained case, the longevity bonds can only be traded at a predetermined lower frequency.

4.3. Assets and liabilities

In this section, we describe the dynamics of the forward interest rate, the forward mortality rates, the bonds and the hedger's assets and liabilities.

4.3.1. Forward interest rates and mortality rates

Interest rates

The instantaneous forward interest rate, $f(t,T)$, is defined as

$$f(t,T) = -\frac{\partial}{\partial T}\log\{B(t,T)\}, \tag{4.8}$$

and the short rate, $r(t)$, is given by $r(t) = f(t,t)$. Under the Q measure, $f(t,T)$ is assumed to follow

$$f(t,T) = f(0,T) + \int_0^t a_f(s,T)ds + \int_0^t \sigma_f(s,T)dW_r^Q(s), \quad t \leq T, \tag{4.9}$$

with a given initial continuous forward curve $f(0,T)$, so that $r(t)$ follows the dynamics

$$r(t) = f(0,t) + \int_0^t a_f(s,t)ds + \int_0^t \sigma_f(s,t)dW_r^Q(s) \tag{4.10}$$

under the Q measure.

Mortality rates

Following Bauer et al. (2008) and Blackburn and Sherris (2013), denote the forward force of mortality as

$$\mu^Q(t,T,x_0,k) = -\frac{\partial}{\partial T}\log\{E_t^Q[{}_{T-t}p(x_0+t,t,k)]\}, \tag{4.11}$$

where $E_t^Q[.] = E^Q[.|\mathcal{F}_t]$ is the time t conditional expectation under the equivalent martingale measure. Given (4.11), we have

$$E_t^Q[{}_{T-t}p(x_0+t,t,k)] = \exp(-\int_t^T \mu^Q(t,s,x_0,k)ds). \tag{4.12}$$

Let $\mu^Q(t,T,x_0) = (\mu^Q(t,T,x_0,rp), \mu^Q(t,T,x_0,pp))'$. $\mu^Q(t,T,x_0)$ is assumed to follow

$$\begin{aligned} \mu^Q(t,T,x_0) =& \mu(0,T,x_0) + \int_0^t a_\mu(s,T,x_0)ds + \int_0^t \sigma_\mu(s,T,x_0)dW_\mu^Q(s), \\ \mu(0,T,x_0) >& 0, \end{aligned} \tag{4.13}$$

where $t \to a_\mu(t,T,x_0)$ and $t \to \sigma_\mu(t,T,x_0)$ are by assumption a continuous and deterministic vector-valued and matrix-valued function, respectively. The vector of the spot force of mortality, $\hat{\mu}^Q(t,x_0)$, is given by

$$\hat{\mu}^Q(t,x_0) \equiv \mu^Q(t,t,x_0). \tag{4.14}$$

Heath et al. (1992) gives a parsimonious dynamic for $a_f(t,T)$ under the equivalent

martingale measure

$$a_f(t,T) = \sigma_f(t,T)\int_t^T \sigma_f(t,s)'ds. \tag{4.15}$$

Following the same argument, we have (see Bauer and Ruß 2006)

$$a_\mu(t,T,x_0) = \sigma_\mu(t,T,x_0)\int_t^T \sigma_\mu(t,s,x_0)'ds. \tag{4.16}$$

Preferably, we should calibrate $f(t,T)$ and $\mu^Q(t,T,x_0)$ from existing data of market prices. However, at the current stage, the price data of longevity-linked instruments is limited, and it is very difficult to obtain meaningful estimations from them. As an alternative, following, for example, Bauer et al. (2008), we model the *best estimated* force of mortality, $\mu(t,T,x_0)$, which can be calibrated from historical mortality data, where the best estimated forward force of mortality (per component) is given by

$$\mu(t,T,x_0,k) = -\frac{\partial}{\partial T}\log\{E_t^P[{}_{T-t}p(x_0+t,t,k)]\}, \quad k = rp, pp. \tag{4.17}$$

where $E_t^P[.] = E^P[.|\mathcal{F}_t]$ is the time t conditional expectation under the physical probability measure. As shown in Bauer et al. (2008), the relation between $\mu(t,T,x_0)$ and $\mu^Q(t,T,x_0)$ is

$$\mu(t,T,x_0) = \mu^Q(t,T,x_0) + \int_t^T \sigma_\mu(s,T,x_0)\lambda_\mu(s)ds. \tag{4.18}$$

From (4.13) and (4.18), the dynamic of $\mu(t,T,x_0)$ is given by

$$\begin{aligned}\mu(t,T,x_0) =& \mu(0,T,x_0) + \int_t^T a_\mu(s,T,x_0)dt + \int_t^T \sigma_\mu(s,T,x_0)dW_\mu^P(t),\\ \mu(0,T,x_0) >& 0.\end{aligned} \tag{4.19}$$

Similarly, define the best estimated spot force of mortality, $\hat{\mu}(t,x_0)$, as

$$\hat{\mu}(t,x_0) \equiv \mu(t,t,x_0). \tag{4.20}$$

In this chapter, we assume that the market price of mortality risk process, $(\lambda_\mu(t))_{t\geq 0}$, is deterministic for mathematical convenience. Consequently, the risk neutral expected survival probabilities can be written as deterministic functions of the best estimated

survival probabilities (Bauer et al. 2008)

$$E_t^Q[{}_{T-t}p(x_0+t,k)] = e^{\int_t^T \int_t^s \sigma_\mu(u,s,x_0)\lambda(u)duds} E_t^P[{}_{T-t}p(x_0+t,k)]. \tag{4.21}$$

As a result, if the values of $(\lambda_\mu(t))_{t\geq 0}$ are given, then we can price the longevity bond and the longevity contingent liabilities using the best estimated forward force of mortality, $\mu(t,T,x_0)$.

4.3.2. Assets and Liabilities

Zero Coupon Bond

For any $T \in \mathcal{T}$, the time t price of the zero coupon bond with any maturity in $\mathcal{T}$ can be written as

$$B(t,T) = \exp(-\int_t^T f(t,s)ds), \tag{4.22}$$

with the dynamics under the equivalent martingale probability measure given by

$$dB(t,T) = B(t,T)r(t)dt + B(t,T)\sigma_B(t,T)dW_r^Q(t), \quad B(T,T) = 1, \tag{4.23}$$

with $\sigma_B(u,T) = -\int_u^T \sigma_f(u,s)ds$. Consequently, the dynamics of the bond under the physical measure is

$$dB(t,T) = B(t,T)(r(t) + b_B(t,T))dt + B(t,T)\sigma_B(t,T)dW_r^P(t), \quad B(T,T) = 1, \tag{4.24}$$

where $b_B(u,T) = -\int_u^T \sigma_f(u,s)ds\lambda_r(u)$, and the bond price in (4.22) can be reformulated as

$$B(t,T) = B(0,T)\exp(\int_0^t [r(s)+b_B(s)-\frac{1}{2}||\sigma_B(s,T)||^2]ds + \int_0^t \sigma_B(s,T)dW_r^P(s)). \tag{4.25}$$

Longevity Bond

A longevity bond contingent on cohort x_0 which matures at T pays ${}_Tp(x_0,0,rp)$ at maturity. Since we consider longevity bonds contingent only on cohort x_0, we simplify the notation by writing $L(t,T) = L(t,T,x_0)$ for all $T \in \tilde{\mathcal{T}}$. The price of the longevity

bond at time t takes the form

$$\begin{aligned} L(t,T) =& B(t,T)E_t^Q[{}_Tp(x_0,0,rp)] \\ =& B(t,T){}_tp(x_0,0,rp)E_t^Q[{}_{T-t}p(x_0+t,t,rp)]. \end{aligned} \tag{4.26}$$

The first equality holds due to the independence of $W_r^Q(t)$ and $W_\mu^Q(t)$. The dynamics of the longevity bond is given by the following Proposition.

Proposition 1. *The price of the zero coupon longevity bond with maturity $T \in \tilde{\mathcal{T}}$ satisfies*

$$\frac{dL(t,T)}{L(t,T)} = \{r(t) + b_B(t,T) + b_L(t,T)\}dt + \sigma_B(t,T)dW_r^P(t) + \sigma_L(t,T)dW_\mu^P(t), \tag{4.27}$$

with $L(T,T) = e^{-\int_0^T \hat{\mu}(s,x_0,rp)ds}$, and with

$$\begin{aligned} \sigma_L(u,T) =& -\int_u^T \sigma_{\mu,1}(u,s,x_0)ds \\ b_B(u,T) =& -(\int_u^T \sigma_f(u,s)ds)\lambda_r(u) \\ b_L(u,T) =& -(\int_u^T \sigma_\mu(u,s,x_0,1)ds)\lambda_\mu(u), \end{aligned} \tag{4.28}$$

where $\sigma_\mu(u,s,x_0,1)$ is the first row of $\sigma_\mu(u,s,x_0)$.

Proof. See Appendix. □

From (4.27), we see that, compared to the dynamics of the zero coupon bond, the longevity bond is in addition affected by the longevity risk premium, $b_L(t,T)$, and the shocks to the mortality processes, $W_\mu^P(t)$.

Pension assets and liabilities

As described in Section 2, the hedger's time t wealth is given by

$$w(t) = u_0(t)B(t) + \boldsymbol{u}_1(t)'\boldsymbol{L}(t) + \boldsymbol{u}_2(t)'\boldsymbol{B}(t). \tag{4.29}$$

For any $\boldsymbol{u}(t) \in \mathcal{U}$, $w(t)$ satisfies the stochastic differential equation

$$dw(t) = \{w(t)r(t) + \boldsymbol{u}(t)'a_w(t)\}dt + \boldsymbol{u}(t)'\sigma_w(t)dW^P(t), \tag{4.30}$$

where $a_w(t)$ is a $\tilde{M} + M$ dimensional vector which contains $L(t,T)(b_B(t,T) + b_L(t,T))$ for $T \in \tilde{\mathcal{T}}$ in the first $\tilde{M}$ components, and $B(t,T)b_B(t,T)$ for $T \in \mathcal{T}$ in the last M component. $\sigma_w(t)$ is the $(\tilde{M}+M)\times(n_r+n_\mu)$ instantaneous covariance matrix containing four blocks

- $\sigma_{LB}(t)$: the upper-left $\tilde{M} \times n_r$ block, which contains $1 \times n_r$ vectors, $L(t,T)\sigma_B(t,T)$, for $T \in \tilde{\mathcal{T}}$.

- $\sigma_B(t)$: the lower-left $M \times n_r$ block, which contains $1 \times n_r$ vectors, $B(t,T)\sigma_B(t,T)$, for $T \in \mathcal{T}$.

- $\sigma_L(t)$: the upper-right $\tilde{M} \times n_\mu$ block, which contains $1 \times n_\mu$ vectors, $L(t,T)\sigma_L(t,T)$, for $T \in \tilde{\mathcal{T}}$.

- the lower-right $\tilde{M} \times n_\mu$ zero sub-matrix.

For simplicity of notation, we denote by $\sigma_w(t,1)$ and $\sigma_w(t,2)$ the $\tilde{M} \times (n_r + n_\mu)$ and $M \times (n_r + n_\mu)$ matrix containing the first $\tilde{M}$ and last M rows of $\sigma_w(t)$, respectively. Moreover, we assume that $\sigma_B(t)$ is invertible for any t. Since we assume that $M = n_r$, this assumption simply means that there is no redundant zero coupon bond in the market.

As discussed in Section 2, the time t discounted market value of pension liabilities, $Y(t)$, can be formulated as

$$\begin{aligned} Y(t) &= \int_{T_0}^{T_1} B(t,s)E_t^Q[{}_sp(x_0,0,pp)]ds \\ &= {}_tp(x_0,0,pp)\int_{T_0}^{T_1} B(t,s)E_t^Q[{}_{s-t}p(x_0+t,t,pp)]ds \\ &= e^{-\int_0^t \hat{\mu}(\tau,x_0,pp)d\tau}\int_{T_0}^{T_1} B(t,s)E_t^Q[{}_{s-t}p(x_0+t,t,pp)]ds. \end{aligned} \tag{4.31}$$

4.4. Benchmark optimization problem

In this section we derive the optimal strategy of the benchmark optimization problem, where all bonds can be traded continuously. First, denote by $X(t, x_0)$ the vector of state variables, which includes the spot interest rate and the spot mortality rate. In a multi-factor setting, i.e., when $n_r > 1$ or $n_\mu > 1$ holds, extra state variables are needed to make the spot interest rate and mortality rates Markovian (see, e.g., Inui and Kijima 1998). In our setup, $X(t, x_0)$ is an $n_r + 2n_\mu$-dimensional vector, containing, as components, the state variables

$$\begin{aligned}
\eta_{i,r}(t) &= \int_0^t \sigma_{i,f}(s,t)\lambda_{i,r}(s)ds + \int_0^t \sigma_{i,f}(s,t)dW_{i,r}^P(s), \quad i = 1,2,3,...,n_r, \\
\eta_{j1,\mu}(t) &= \int_0^t \sigma_{j,\mu}(s,t,x_0,rp)dW_{j,\mu}^P(s), \quad j = 1,2,3,...,n_\mu, \\
\eta_{j2,\mu}(t) &= \int_0^t \sigma_{j,\mu}(s,t,x_0,pp)dW_{j,\mu}^P(s), \quad j = 1,2,3,...,n_\mu,
\end{aligned} \tag{4.32}$$

where $\sigma_{i,f}(s,t)$, $\lambda_{i,r}(s)$ and $\sigma_{j,\mu}(s,t,x_0,k)$ are the i-th entry of $\sigma_f(s,t)$ and $\lambda_r(s)$, and the j-th entry of $\sigma_\mu(s,t,x_0,k)$, $k = rp, pp$, respectively. In this section, we simply write the dynamics of $X(t, x_0)$ under the physical measure P as

$$dX(t,x_0) = a_X(t, X_t, x_0)dt + \sigma_X(t, X_t, x_0)dW^P(t). \tag{4.33}$$

The concrete representation of the dynamics can be derived when specific parametric choices of $\sigma_f(s,t)$ and $\sigma_\mu(s,t,x_0)$ are made. An illustration is provided in the next section. For simplicity of notation, we omit the term x_0 in the drift and volatility terms of the state variables in the sequel.

We first consider the case where the hedger is able to continuously rebalance her position in both the zero coupon bond and the longevity bond. The hedger solves the problem

$$\min_{\boldsymbol{u}\in\mathcal{U}} \mathrm{Var}_0[e^{-\int_0^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))], \tag{4.34}$$

under the budget constraint (4.6). As stated in Basak and Chabakauri (2012), most of the existing literature only characterize the pre-commitment dynamic hedging strategy which minimizes the variance at the initial date. We consider the optimal (time-consistent)

strategy when the hedger would also minimize the variance at later dates. Applying the law of total variance to the variance criterion evaluated at time t yields

$$\begin{aligned}
&\mathrm{Var}_t[e^{-\int_t^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))] \\
=&E_t[\mathrm{Var}_{t+\epsilon}[e^{-\int_{t+\epsilon}^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))]] + \mathrm{Var}_t[E_{t+\epsilon}[e^{-\int_t^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))] \\
\neq&E_t[\mathrm{Var}_{t+\epsilon}[e^{-\int_{t+\epsilon}^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))]]. \qquad (4.35)
\end{aligned}$$

Intuitively speaking, the time $t+\epsilon$ variance of the time T_0 hedging error is smaller than the time t variance. Being at time t, the hedger would choose a hedging strategy that not only accounts for the expected time $t+\epsilon$ variance of the time T_0 hedging error, but also for the variance of time $t+\epsilon$ expected hedging error. However, the latter term is not included anymore in the objective function as the time interval ϵ elapses, and the hedger may deviate from the time t optimal strategy at time $t+\epsilon$, if she minimizes the variance at time $t+\epsilon$. In this chapter, we follow the recursive approach proposed by Basak and Chabakauri (2010, 2012), Strotz (1956), and Caplin and Leahy (2006) to obtain time-consistent optimal solutions to the minimum-variance problem. In particular, the recursive formulation at time t is expressed as the expected future value of the variance plus an adjustment term, which is the time t variance of the expected terminal net asset value. For each $t \in [0, T_0]$, define

$$U_t \equiv \mathrm{Var}_t[-e^{-\int_t^{T_0} r(\tau)d\tau}(Y(T_0) - w(T_0))]. \qquad (4.36)$$

Substituting the second line in (4.35) into (4.36) yields

$$U_t = E_t[U_{t+\epsilon}] + \mathrm{Var}_t[E_{t+\epsilon}[-e^{-\int_t^{T_0} r(\tau)d\tau}(Y(T_0) - w(T_0))]]. \qquad (4.37)$$

The hedger minimizes (4.37) subject to the budget constraint (4.6) and (4.33) by backward induction, yielding a time-consistent hedging strategy. In particular, the optimal hedging strategy, $u^*(.)$, is given by Theorem 1.

Theorem 1. *Under the budget constraints* (4.6) *and* (4.33)*, the minimum-variance optimal strategy is given by*

$$\boldsymbol{u}^*(t) = -(\sigma_w(t)\sigma_w(t)')^{-1}\sigma_w(t)\sigma_X(t)'\frac{\partial G(t)}{\partial X(t)}, \qquad (4.38)$$

where $G(t)$ has the representation[48]

$$G(t) = -E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)]. \tag{4.39}$$

$\tilde{Q}$ is a probability measure with the Radon-Nikodym density w.r.t. P

$$(\frac{d\tilde{Q}}{dP})_t = \exp(-\frac{1}{2}\int_0^t ||\lambda^{\tilde{Q}}(s)||^2 ds - \int_0^t \lambda^{\tilde{Q}}(s) dW^P(s)), \tag{4.40}$$

with

$$\lambda^{\tilde{Q}}(t) = -a_w(t)'(\sigma_w(t)\sigma_w(t)')^{-1}\sigma_w(t). \tag{4.41}$$

The corresponding variance of the hedging error is given by

$$\begin{aligned}
&Var_t[e^{-\int_t^{T_0} r(\tau)d\tau}(Y(T_0) - w^*(T_0))] \\
=&E_t[\int_t^{T_0} e^{-2\int_t^s r(\tau)d\tau}(\frac{\partial G(s)}{\partial X(s)}'\sigma_X(s)\sigma_X(s)'\frac{\partial G(s)}{\partial X(s)} \\
&-\frac{\partial G(s)}{\partial X(s)}'\sigma_X(s)\sigma_w(s)'(\sigma_w(s)\sigma_w(s)')^{-1}\sigma_w(s)\sigma_X(s)'\frac{\partial G(s)}{\partial X(s)})ds],
\end{aligned} \tag{4.42}$$

where $w^*(T_0)$ is the T_0 wealth under the optimal hedging strategy.

Proof. The proof follows directly from Proposition 3.1 and 3.2 of Wong et al. (2014), with the risk aversion parameter ϕ set to 0. In our case, we have

$$\Gamma(t) \equiv G(t) = -E_t[e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0) - \int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \boldsymbol{u}^*(s) a_w(s) ds],$$

and $Y(t) \equiv w(t)$. □

Without loss of generality, we take a further look at the measure $\tilde{Q}$ given in Theorem 1 for $\mathcal{T}_M = \mathcal{T}_{\tilde{M}} = \{\tau\}$, i.e., when there is one zero coupon bond and longevity bond tradable in the market. From the definitions of a_w and σ_w, we have

$$\begin{aligned}
&\lambda^{\tilde{Q}}(s) = -a_w(t)'(\sigma_w(t)\sigma_w(t)')^{-1}\sigma_w(t) \\
=&\Big(\frac{\int_t^T \sigma_f(t,s)ds b_B(t,T)}{\int_t^T \sigma_f(t,s)ds' \int_t^T \sigma_f(t,s)ds}, \quad \frac{\int_t^T \sigma_{\mu,1}(t,s,x_0)ds b_L(t,T)}{\int_t^T \sigma_\mu(t,s,x_0,1)ds' \int_t^T \sigma_\mu(t,s,x_0,1)ds}\Big)'.
\end{aligned} \tag{4.43}$$

Eq (4.43) shows that the measure $\tilde{Q}$ is completely characterized by the price of risk of the

[48]For notational simplicity, we write $G(t, X(t))$ as $G(t)$.

tradable assets, i.e., the zero coupon bond and the zero coupon longevity bond. Although the risk neutral measure is not unique due to market incompleteness, the measure $\tilde{Q}$ can be uniquely determined from observable market prices. Moreover, the optimal strategy given in (4.38) preserves the structure of that in complete markets: when the market would be complete, i.e., there would exist a unique risk neutral measure $\bar{Q}$, the optimal hedging strategy given by the standard no-arbitrage method preserves the same structure as given in (4.38), with the term $G(t)$ given by

$$G(t) = -E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)] = -B(t, T_1)Y(t). \tag{4.44}$$

In other words, $G(t)$ is the negative unique no-arbitrage discounted value of $Y(T_0)$ in this case, and the hedging error, $e^{-\int_t^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))$, becomes 0 for all t. Therefore, we see that the optimal hedging strategy given in Theorem 1 is a simple generalization of the perfect hedge in the complete market, and minimizes the hedging error when the market is incomplete. In fact, $G(t)$ can be written as

$$G(t) = E_t[e^{-\int_t^{T_0} r(\tau)d\tau} w^*(T_0) - w(t)] - E_t[e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)], \tag{4.45}$$

where the first term represents the expected trading gains in the assets that the hedger gives up for hedging the non-tradable liabilities during $[t, T_0]$. For this reason, the measure $\tilde{Q}$ is called the "hedge-neutral" measure by Basak and Chabakauri (2012).

4.5. Optimal strategy under trading constraint

As mentioned above, the longevity-linked capital market is newly developed, and it might not be possible for the hedger to trade longevity-linked instruments at the same high frequency as other financial assets. To cope with this fact, we study the hedging strategies under a liquidity constraint, under which the hedger can only trade the longevity bond at fixed and deterministic times $t \in \{t_0, t_1, ..., t_n\}$. However, following Dahl et al. (2011) and Ang et al. (2014), we still assume that $\boldsymbol{B}(t)$ and $\boldsymbol{L}(t)$ are $\mathcal{F}$-adapted, i.e., the price of both the zero coupon bonds and the longevity bonds are observed by the hedger at any time t, not only at the discrete times t_i, $i = 1, 2, ..., n$. Moreover, the holds of all

assets are allowed to jump. Assume that the trading opportunity arrives at time t, then the hedger is able to rebalance u_0 and u_2 such that

$$0 = (u_0(t) - u_0(t-))B(t) + (\boldsymbol{u}_1(t) - \boldsymbol{u}_1(t-))'\boldsymbol{L}(t) + (\boldsymbol{u}_2(t) - \boldsymbol{u}_2(t-))'\boldsymbol{B}(t) \tag{4.46}$$

holds. Therefore, the budget constraint given in Eq (4.6) still implies self-financing in the constraint case. A similar constrained optimization problem is considered in Dahl et al. (2011), who consider a quadratic loss function.

4.5.1. The constrained optimal strategy

Denote by $\hat{\boldsymbol{u}}(.) = (\hat{\boldsymbol{u}}_1(.)', \hat{\boldsymbol{u}}_2(.)')'$ the hedging strategy under the liquidity constraint, and $\hat{\mathcal{U}}$ the corresponding admissible set. In this case the hedger solves the optimization problem

$$\min_{\hat{\boldsymbol{u}} \in \hat{\mathcal{U}}} \mathrm{Var}_0[e^{-\int_0^{T_0} r(\tau)d\tau}[w(T_0) - Y(T_0)]] \tag{4.47}$$

under the constraints (4.6) and (4.33). Denote by $\sigma_w(s,1)$ and $\sigma_w(s,2)$ the $\tilde{M} \times (n_r + n_\mu)$ and $M \times (n_r + n_\mu)$ matrix containing the upper $\tilde{M}$ and lower M rows of $\sigma_w(s)$, respectively. Moreover, denote by $a_w(s,1)$ and $a_w(s,2)$ the $\tilde{M}$ and M vector containing the first $\tilde{M}$ and last M components of $a_w(s)$. The constrained optimal hedging strategy is given in the following Theorem.

Theorem 2. *Denote $G(t)$ by*

$$\begin{aligned} G(t) = &\sum_{j=i+1}^{n} E_t^{\bar{Q}}[\int_{t_j}^{t_{j+1}} e^{-\int_t^s r(\tau)d\tau} \hat{\boldsymbol{u}}_1^*(t_j) a_w(s,1) ds] - E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau} Y_{T_0}] \\ &+ \hat{\boldsymbol{u}}_1^*(t_i)' E_t^{\bar{Q}}[\int_t^{t_{i+1}} e^{-\int_t^s r(\tau)d\tau} a_w(s,1) ds] \\ \equiv &G_i(t) + \hat{\boldsymbol{u}}_1(t_i)' H_i(t), \end{aligned} \tag{4.48}$$

for $t \in [t_i, t_{i+1})$ with $t_{n+1} \equiv T_0$, and $\bar{Q}$ the probability measure with Radon-Nikodym density w.r.t. P given by

$$(\frac{d\bar{Q}}{dP})_t = \exp(-\frac{1}{2}\int_0^t ||\lambda^{\bar{Q}}(s)||^2 ds - \int_0^t \lambda^{\bar{Q}}(s) dW^P(s)), \tag{4.49}$$

with

$$\lambda^{\bar{Q}}(t) = -a_w(t,2)'(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2). \tag{4.50}$$

Moreover, denote the matrices $A_i(s)$ *and* $B_i(s)$ *by*

$$\begin{aligned} A_i(s) =&\sigma_w(s,1)(I-\bar{\sigma}_w(s,2))\sigma_w(s,1)' + 2\sigma_w(s,1)(I-\bar{\sigma}_w(s,2))\sigma_X(s)'\frac{\partial H_i(s)}{\partial X(s)} \\ &+\frac{\partial H_i(s)}{\partial X(s)}'\sigma_X(s)(I-\bar{\sigma}_w(s,2))\sigma_X(s)'\frac{\partial H_i(s)}{\partial X(s)} \\ B_i(s) =&-(\sigma_w(s,1)+\frac{\partial H_i(s)}{\partial X(s)}'\sigma_X(s))(I-\bar{\sigma}_w(s,2))\frac{\partial G_i(s)}{\partial X(s)} \end{aligned} \tag{4.51}$$

for $s \in [t_i, t_{i+1})$, $i = 1, 2, \cdots, n$. *Assume that* $E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} A_i(s)ds]$ *is invertible for all* i. *Under the budget constraints* (4.29) *and* (4.33), *the minimum-variance optimal strategy to the problem* (4.47) *is given by*

$$\begin{aligned} \hat{\boldsymbol{u}}_1^*(t) =&(E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} A_i(s)ds])^{-1}E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} B_i(s)ds], \quad t \in [t_i, t_{i+1}). \\ \hat{\boldsymbol{u}}_2^*(t) =&-(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_X(t)'\frac{\partial G(t)}{\partial X(t)} \\ &-(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_w(t,1)'\boldsymbol{u}_1^*(t_i), \quad t \in [t_i, t_{i+1}) \end{aligned} \tag{4.52}$$

for $0 \le i \le n$.

Proof. See Appendix. □

The invertibility assumption in Theorem 2 can be satisfied under reasonable parametrization of the interest rate and mortality rate processes. An illustrating parametric specification is given in the next section.

4.5.2. Comparison with the benchmark case

In this subsection, we compare the constrained optimal strategy in (4.52) with the benchmark optimal strategy given in (4.38).

Optimal holding of zero coupon bonds

The benchmark optimal $\boldsymbol{u}_2^*$ given in (4.38) can be formulated as

$$\begin{aligned}\boldsymbol{u}_2^*(t) =& -(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_X(t)'\frac{\partial G(t)}{\partial X(t)}\\ &-(\sigma_w(t,2)\sigma_w(t,1)')^{-1}\sigma_w(t,2)\sigma_w(t,2)'\boldsymbol{u}_1^*(t),\end{aligned} \tag{4.53}$$

with the $G(t)$ given in (4.39). Therefore, the structure of the optimal holding of zero coupon bonds is similar in both cases. In fact, comparing (4.52) and (4.53), we see that, for any given $\boldsymbol{u}_1$, the optimal holding of the zero coupon bond hedges the interest rate risk of two parts: the market value of the liabilities (the first part in (4.53)) under the hedge neutral measure, and the holding of longevity bonds (the second part in (4.53)). However, the hedge neutral measure also depends on $\boldsymbol{u}_1$. For the benchmark case, we can write $G(t)$ in (4.39) as

$$G(t,\boldsymbol{u}_1) = E_t^{\bar{Q}}[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau}\boldsymbol{u}_1(s)'a_w(s,1)ds] - E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}Y(T_0)], \tag{4.54}$$

with $\bar{Q}$ defined in (4.49). When evaluated at the optimal $\boldsymbol{u}_1(t)$, $G(t,\boldsymbol{u}_1(t))$ becomes (By Feynman-Kac Theorem)

$$\begin{aligned}G(t,\boldsymbol{u}_1^*) =& E_t^{\bar{Q}}[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau}\boldsymbol{u}_1^*(t)'a_w(s,1)ds] - E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}Y(T_0)]\\ =& -E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}Y(T_0)],\end{aligned} \tag{4.55}$$

with $\tilde{Q}$ defined in Theorem 1. Similarly, for the constrained case, we can see from (4.48) that the constrained hedge neutral measure, denoted by $\tilde{Q}_c$, should satisfy the relation

$$\begin{aligned}E_t^{\tilde{Q}_c}[e^{-\int_t^{T_0} r(\tau)d\tau}Y(T_0)] =& \sum_{j=i+1}^{n} E_t^{\bar{Q}}[\int_{t_j}^{t_{j+1}} e^{-\int_t^s r(\tau)d\tau}\hat{\boldsymbol{u}}_1^*(t_j)a_w(s,1)ds] - E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}Y(T_0)]\\ &+\hat{\boldsymbol{u}}_1^*(t_i)'E_t^{\bar{Q}}[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau}a_w(s,1)ds]\end{aligned} \tag{4.56}$$

for $i = 1, 2, ..., n$. From the above analysis, we see that the holding of the longevity bond affects the corresponding optimal holding of the zero coupon bond via two channels: a direct channel by introducing extra interest rate risk; and an indirect channel by changing

the hedge neutral measure.

Optimal holding of longevity bonds

The benchmark optimal $\boldsymbol{u}_1$ given in (4.38) can be formulated as

$$\boldsymbol{u}_1^*(t) = (\tilde{A}(t))^{-1}\tilde{B}(t), \tag{4.57}$$

with

$$\begin{aligned}\tilde{A}(t) =&\sigma_w(s,1)(I-\bar{\sigma}_w(s,2))\sigma_w(s,1)'\\ \tilde{B}(t) =&-\sigma_w(s,1)(I-\bar{\sigma}_w(s,2))\sigma_X(s)\frac{\partial G(s)}{\partial X(s)}.\end{aligned} \tag{4.58}$$

Compare (4.57) with (4.52), we see that the structure of the optimal holding of longevity bonds is similar in both cases. Intuitively speaking, the benchmark optimal $\boldsymbol{u}_1$ minimizes the portfolio's *instantaneous* sensitivity with respect to the mortality risk under the *benchmark hedge neutral measure*, while the constrained optimal $\hat{\boldsymbol{u}}_1$ minimizes the portfolio's *expected accumulated* sensitivity with respect to the mortality risk in each $[t_i, t_{i+1})$ under the *constrained hedge neutral measure*. The difference between the benchmark $\tilde{A}$ and $\tilde{B}$ matrix and the constrained A and B matrix given in (4.51) comes from the difference in the hedge neutral measure. In particular, in both cases, the hedge neutral measure at time t depends on future optimal $\hat{\boldsymbol{u}}_1$. In the benchmark case, the future optimal $\hat{\boldsymbol{u}}_1$ is independent of the optimal $\hat{\boldsymbol{u}}_1$ at time t. However, for the constrained case, for every $t \in [t_i, t_{i+1})$, the future optimal $\hat{\boldsymbol{u}}_1$ up to time t_{i+1} depends on the current optimal $\hat{\boldsymbol{u}}_1$.

We have derived the optimal strategy under the trading constraint of longevity bonds. However, from (4.51) – (4.52), we see that, for $i \leq n-1$, the constrained optimal $\hat{\boldsymbol{u}}_1$ at $t \in [t_i, t_{i+1})$ depends explicitly on the future optimal $\hat{\boldsymbol{u}}_1$ (through the term $G_i(t)$). As a result, in the application, we need to solve the optimal $\hat{\boldsymbol{u}}_1$ period by period. However, the optimal $\hat{\boldsymbol{u}}_1$ in each period is a rather complicated expression, which makes the application inconvenient. As an alternative, we consider the optimization problem under an additional constraint, which substantially increase the tractability of the optimal strategies. We will see in the next section that satisfying performance of the optimal hedging

strategy can be achieved under this additional constraint.

4.5.3. The constrained optimal strategy in a special case

In this subsection, we consider the constrained optimization problem under an additional constraint

$$E_t[W^*(T_0)] = E_t[W(T_0, \hat{\boldsymbol{u}})], \;\; t \in [0, T_0], \tag{4.59}$$

where $W^*(T_0)$ is the T_0 wealth under the benchmark optimal strategy given in (4.38). Constraint (4.59) means that the hedger only considers the set of constrained strategies that give the same expected T_0 market value of investments as the benchmark optimal strategy. The optimal strategy under this additional constraint is given by the next Proposition.

Proposition 2. *Under the budget constraints* (4.29), (4.33) *and* (4.59), *the minimum-variance optimal strategy to the problem* (4.47) *is given by*

$$\begin{aligned}
\hat{\boldsymbol{u}}_1^*(t) =& (E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} A_i(s)ds])^{-1} E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} B_i(s)ds], \;\; t \in [t_i, t_{i+1}). \\
\hat{\boldsymbol{u}}_2^*(t) =& -(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_X(t)' \frac{\partial G(t)}{\partial X(t)} \\
& -(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_w(t,1)'\boldsymbol{u}_1^*(t_i), \;\; t \in [t_i, t_{i+1})
\end{aligned} \tag{4.60}$$

under the assumption that $E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^{s} r(\tau)d\tau} A(s,t_i)ds]$ *is invertible for any* i. *The matrices* $A(s,t_i)$ *and* $B(s,t_i)$ *are given by*

$$\begin{aligned}
A(s,t_i) =& \sigma_w(s,1)(I - \bar{\sigma}_w(s,2))\sigma_w(s,1)' \\
B(s,t_i) =& -\sigma_w(s,1)(I - \bar{\sigma}_w(s,2))\sigma_X(s)\frac{\partial G(s)}{\partial X(s)}
\end{aligned} \tag{4.61}$$

for all i. *The* $G(t)$ *is given in* (4.39).

Proof. The proof follows Theorem 2 directly, with the $G_i(t)$ replaced by $G(t)$ for every i. □

Compare (4.61) with (4.58), we see that the constrained optimal $\hat{\boldsymbol{u}}_1$ in this case looks more similar to the benchmark optimal $\hat{\boldsymbol{u}}_1$. Specifically, the constrained optimal $\hat{\boldsymbol{u}}_1$ now only minimizes the portfolio's expected accumulated sensitivity with respect to the

mortality risk, without changing the hedge neutral measure. Indeed, under the additional constraint (4.59), we have

$$\begin{aligned} G(i,t) =& E_t[\int_{t_j}^{t_{j+1}} e^{-\int_t^s r(\tau)d\tau} \hat{\boldsymbol{u}}^*(s) a_w(s) ds] - E_t[e^{-\int_t^{T_0} r(\tau)d\tau} Y_{T_0}] \\ =& E_t[\int_{t_j}^{t_{j+1}} e^{-\int_t^s r(\tau)d\tau} \boldsymbol{u}^*(s) a_w(s) ds] - E_t[e^{-\int_t^{T_0} r(\tau)d\tau} Y_{T_0}] \\ =& G(t). \end{aligned} \tag{4.62}$$

In other words, under the additional constraint (4.59), the constrained optimal strategy depends on the benchmark hedge neutral measure, and thus the optimal $\hat{\boldsymbol{u}}_1$ does not depend on future $\hat{\boldsymbol{u}}_1$-s anymore.

4.6. Numerical evaluation of the optimal hedging strategies

In Section 4 and 5 we derived optimal hedging strategies using a general HJM framework. In this section we give a numerical illustration of the optimization problem considered in Section 4.5.3 using a specific parametrization of the forward interest rate and mortality rate processes. Moreover, we evaluate the hedging effectiveness using parameter estimates from existing literature.

Parametrization

As the focus of this chapter is the hedging of longevity risk, we impose a simplified structure for the interest rates. In particular, we define

$$\sigma_f(t,T) = \beta e^{\kappa(T-t)}, \tag{4.63}$$

i.e., we consider a one factor Hull-White specification of the interest rate process. For the mortality process, we define

$$\sigma_\mu(t,T,x_0) = \begin{pmatrix} c_{11}(x_0)e^{\omega_{11}(T-t)} & c_{12}(x_0)e^{\omega_{12}(T-t)} & c_{13}(x_0)e^{\omega_{13}(T-t)} \\ c_{21}(x_0)e^{\omega_{21}(T-t)} & c_{22}(x_0)e^{\omega_{22}(T-t)} & c_{23}(x_0)e^{\omega_{23}(T-t)} \end{pmatrix}, \tag{4.64}$$

where the $c_{kj}(x_0) = c_{kj}e^{a_{kj}x_0}$. c_{kj}-s, a_{kj}-s, and ω_{kj}-s are parameters which need to be determined. The derivation of u^* in (4.38) and (4.52) under the specifications (4.63) and (4.64) is rather lengthy, and is left to the Appendix.

Choice of parameter values

For the one factor Hull-White interest rate model, we use the estimation results from Driessen et al. (2003), who estimate a one-factor Hull-White model from U.S. interest rate data: $\beta = 0.0095$ and $\kappa = -0.009$. For the mortality model, we use the estimation results of the dependent factor model in Blackburn and Sherris (2014), who apply a three-factor two population model to Australian and Swedish males data. The parameter values are given by

$$(\omega_{11}, \omega_{12}, \omega_{13}) = (\omega_{21}, \omega_{22}, \omega_{23}) = (0.1246, 0.08366, 0.1714), \tag{4.65}$$

and

$$\begin{aligned}(c_{11}, c_{12}, c_{13}) =&(0.00006657, 0.0002238, 0.000001095)\\ (c_{21}, c_{22}, c_{23}) =&(0.00003435, 0.0003469, 0.000001095).\end{aligned} \tag{4.66}$$

The third factor has the same effect to both populations, and the time sensitivity parameters, ω_{ij}-s, are population neutral.

Besides the specified parameter values, initial forward curves are also needed to generate future interest and mortality rates. For the interest rate, we generate the forward curve based on the yield curve on September, 30, 2014, reported by the U.S. Department of the treasury.[49] For the mortality rates, we generate the initial forward curve using the Lee-Carter (1992) model and mortality data downloaded from the Human Mortality Database.[50] The age groups and sample period chosen to generate the initial forward curve are the ages 21 to 100 and the years 1965 to 2009, respectively. Finally, we use a constant vector of price of risk: $\lambda = (0.05333, 0.3008, 0.2898, 0.2788)'$.[51]

[49] http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield.

[50] http://www.mortality.org/.

[51] The price of longevity risk is calibrated in Bauer et al. (2010). For the price of interest risk, De Jong and Santa-Clara (1999) propose a square root price of risk process, $\lambda_r(t) = \bar{\lambda}\sqrt{\sigma_0 + \sigma_1 r(t)}$, and calibrate the parameter values: $\bar{\lambda} = 47.86$, $\sigma_0 = 0$, and $\sigma_1 = 0.0579$. We impose a constant price of risk which equals to the mean of the $\lambda_r(t)$ process generated using their parameters, i.e., $\lambda_r = \overline{\lambda_r(t)}$.

Strategy	Benchmark	Constrained	Static	Interest Only
Zero Coupon Bond	week	week	week	week
Longevity Bond	week	1/2/5 year	buy and hold	no trading

Table 4.1: The trading frequency of zero coupon bond and longevity bond for the four strategies.

Optimal strategies

For illustration purpose, we consider the case where the hedger trades only one zero coupon bond and one longevity bond. In particular, we assume that $\mathcal{T} = \tilde{\mathcal{T}} = \{T_1\}$. In other words, both bonds mature at the same date when the last possible payment in the hedger's liabilities is made. We evaluate the performance of four strategies: the benchmark strategy, a constrained strategy, a static strategy, and an interest-only strategy. For each strategy, the zero coupon bond is assumed to be trade weekly, while the longevity bond is assumed to have different trading frequencies, see Table 4.1.

The interest-only strategy, where the hedger does not trade the longevity bond at all, is a special case of the benchmark strategy. The optimal interest-only strategy is given by the next proposition.

Proposition 3. *The optimal strategy and the value function in the case where the hedger only trades zero coupon bonds have the same forms as in* (4.38) *and* (4.42) *with* $u_1(t) = 0$ *and* $\sigma_w(t) = (B(t,T)\sigma_B(t,T), 0_{n_\mu})$ *for* $t \in [0, T_0]$.

Proof. The proof follows directly from the proof of Theorem 1. □

For the portfolio, we consider payments to the cohort that is aged 55 in the year 2009 and start 10 years from now. In other words, we let $x_0 = 55$ and $T_0 = 10$. Moreover, we let $T_1 = 30$, i.e., the last possible payment is made when the cohort reaches age 85. Figure 4.2 reports the best estimated survival probabilities of cohort x_0 for both populations. We see that the estimated survival probabilities are different for each population, which shows the presence of the population basis risk.

We generate 1000 pathes of the state variables, and compute corresponding realizations of the four optimal strategies.[52] The mean optimal strategies for the first three cases are reported in Figure 4.3, where the left column displays the optimal holdings

[52]With 1000 realizations, the mean, 1% and 99% quantiles of the time 0 optimal standard deviation of the hedger's hedging error under the benchmark strategy is 0.0160, 0.0154 and 0.0168. Confidence intervals for the optimal standard deviations under other strategies are also small.

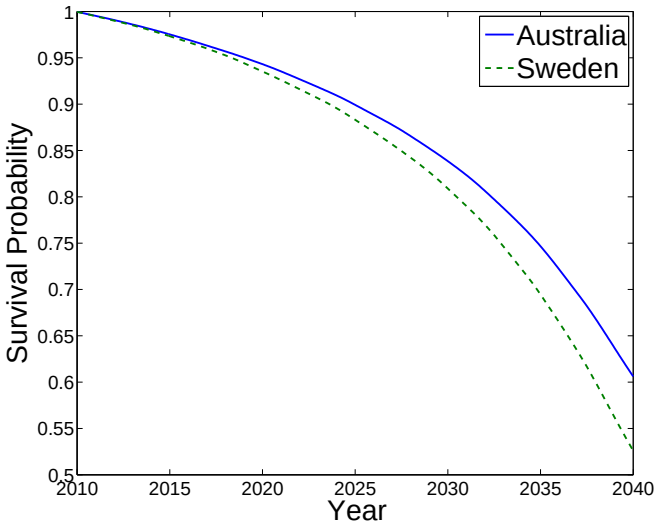
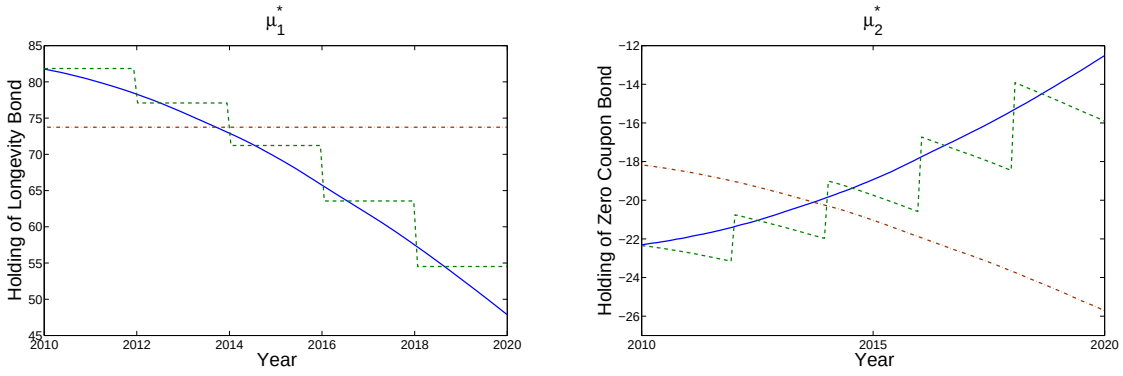


Figure 4.2: The survival probability for the males cohort aged 55 at year 2009. The solid line and dashed line is the survival probability for the Australian and the Swedish population, respectively.

of the longevity bond, and the right column reports the optimal holding of the zero coupon bond. For the constrained case, we only report the optimal holdings for the 2-year frequency, since the patterns for the 1- and 5-year-frequency strategy are very similar. In each figure, the solid line, the dashed line, and the dotted line represents the optimal strategy for the benchmark case, the constrained case, and the static case, respectively. Moreover, the mean optimal strategy, as well as the 5% and 95% quantiles, in the interest-only case is reported in Figure 4.4. We see that, firstly, for the two dynamic strategies, the optimal holding of the longevity bond is decreasing in time. This observation is intuitive, since as the dates of payment become close, the mortality rates affecting the liabilities become less uncertain. As a result, the longevity risk exposure in the liabilities is decreasing over time, and the hedger would gradually reduce her holding of the longevity bond.

For the zero coupon bond, the patterns are more complicated. The optimal holding of the zero coupon bond is increasing in the benchmark case, decreasing in the static case and the interest-only case, and has a decreasing trend with upward jumps in the constrained case. These observations result from the interaction of two opposite forces. First, the duration of the liabilities decreases over time, and so does the interest rate exposure in the liabilities. This effect pushes the hedger to reduce her holding of zero-coupon bonds. As can be seen most clearly in the interest-only case, the hedger initially

holds a positive amount of zero-coupon bond to hedge her interest rate risk, and then gradually reduces her holding. Second, as mentioned in Section 4, the price of the zero coupon bond and the longevity bond move in the same direction as interest rate changes. Therefore, the holding of the longevity bond introduces extra interest rate risk, which needs to be hedged by holding zero-coupon bond in the opposite direction. As can be seen in the first three cases, the large holding of the longevity bonds at time 0 introduces extra interest rate risk, which requires the hedger to short the zero coupon bond. If the holding of the longevity bonds is unchanged over time, as in the static case and the intervals in the constrained case, the hedger would even short more zero-coupon bonds as the duration of her liabilities decreases. However, if the holding of the longevity bond decreases over time, as in the benchmark case and the trading points of the longevity bond in the constrained case, the hedger would reduce her holdings (in absolute term) of the zero-coupon bond correspondingly.



(a) Longevity bond (b) Zero coupon bond

Figure 4.3: The mean optimal strategies for the benchmark case, the constrained case (2-year frequency), and the static case. The solid line, the dashed line, and the dotted line represent the optimal strategy in the benchmark case, the constrained case, and the static case, respectively.

Besides the mean optimal strategies, we also investigate how sensitive the optimal strategies are to random realizations of the state variables. The optimal strategies turn out to be with a clear monotonic trend, except for the optimal holding of the zero-coupon bond in the constrained case. However, even in this case, the optimal holding is monotonically decreasing between each two consecutive trading points. Therefore, we examine the sensitivity of the strategies by looking at their confidence intervals. The 5% and 95% quantiles of the benchmark, the constrained, and the static optimal strategies are reported from top to bottom in Figure 4.5. In each row, the left column reports

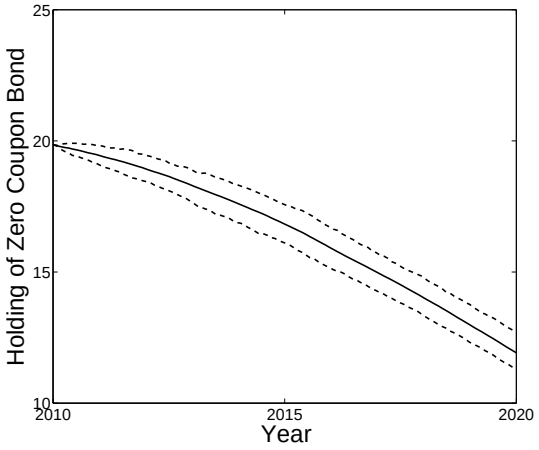


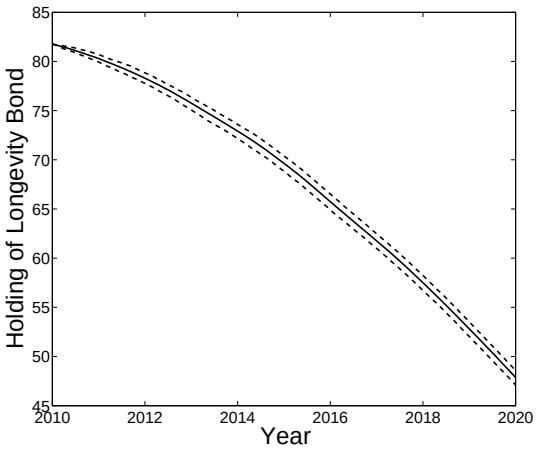
Figure 4.4: The mean, 5%, and 95% quantiles of the optimal holding of the zero-coupon bond in the interest-only case.

the optimal holdings of the longevity bond, and the right column reports the optimal holdings of the zero-coupon bond. Together with Figure 4.4, we see that the confidence intervals of the optimal strategies in all cases are small. The possible reason is that, with the parameter estimates we use, the volatilities of the interest rate and mortality rate forecast are small.

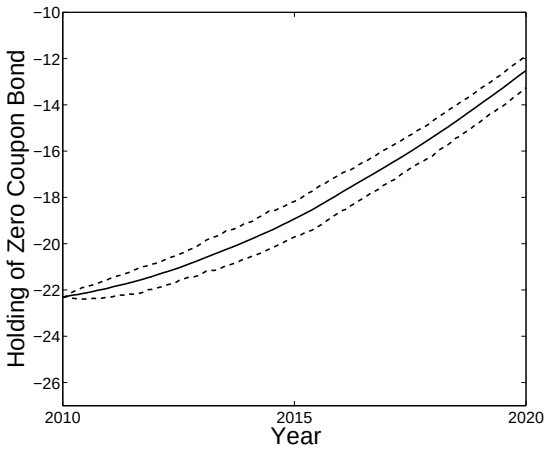
Finally, we evaluate the performance of different hedging strategies. In particular, the time 0 standard deviation (std.) of the hedger's hedging error under the benchmark strategy is 0.0160. This number can be interpreted using the following example. Assume that a male receives annual pension payment of 46,000 U.S. dollars after he retires,[53] then the time 0 std. of the hedger's hedging error corresponding to one pension member is 736 U.S. dollars, given that she follows the benchmark hedging strategy. In Table 4.2 we report the ratio of the time 0 std. under the other strategies to the std. of the benchmark strategy. Continuing the above example, the time 0 std. would be about 3.7% and 18% higher if she could only trade the longevity bond on a 2-year or 5-year frequency, respectively. Moreover, if a hedger switches from static hedging (10-year frequency) to 5-year frequency, the optimal std. would be reduced by around 23%. Finally, the std. would be about 7.44 times higher if the hedger does not hedge longevity risk at all.

From the numerical study, we see that lowering the trading frequency of the longevity bond from weekly to a 2-year frequency only leads to a slight decrease of the hedging quality. However, compared with dynamic hedging strategies, even the constrained one, a

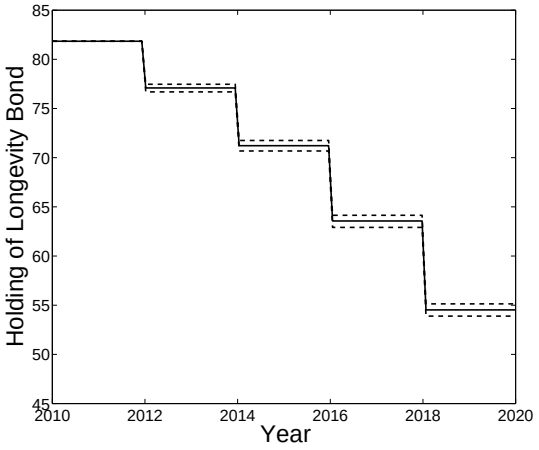
[53]The number is calculated from OECD (2013).



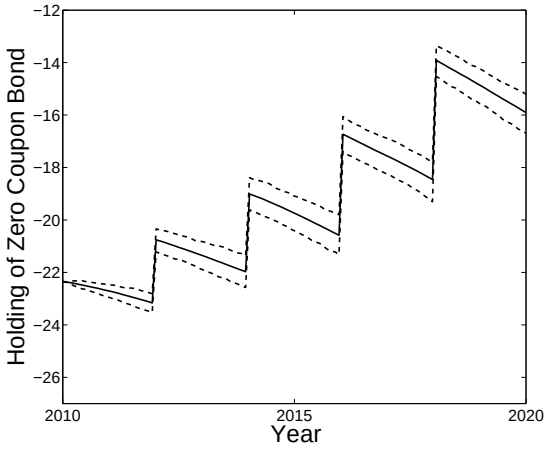
(a) u_1^*



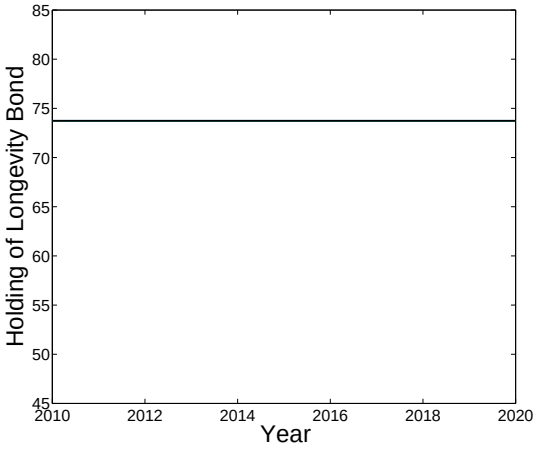
(b) u_2^*



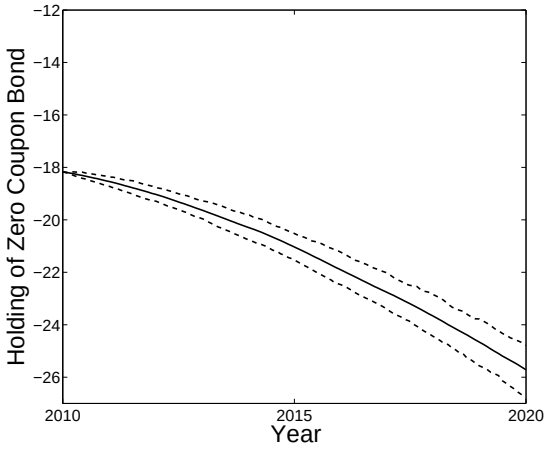
(c) u_1^*



(d) u_2^*



(e) u_1^*



(f) u_2^*

Figure 4.5: The mean, 5%, and 95% quantiles of the optimal strategies for the benchmark case, the constrained case (2-year frequency), and the static case (from top to bottom). In each row, the left figure reports the optimal holding of the longevity bond, and the right figure reports the optimal holding for the zero-coupon bond.

	1-year	2-year	5-year	Static	Interest-only
Ratio of the std.	101.09%	103.7%	118.25%	153.54%	743.9%

Table 4.2: The ratio of the time 0 std. of the optimal hedging error relative to the std. of the benchmark case. In each case, the ratio equals $(\mathrm{Var}_t[e^{-\int_0^{T_0} r(\tau)d\tau}(Y(T_0) - w_s^*(T_0))]/\mathrm{Var}_t[e^{-\int_0^{T_0} r(\tau)d\tau}(w_b^*(T_0) - Y(T_0))])$, where $w_b^*(T_0)$ and $w_s^*(T_0)$ are the time T_0 assets under the benchmark strategy and the corresponding strategy, respectively.

	Benchmark	1-year	2-year	5-year	Static	Interest-only
W. p. risk	0.0160	101.09%	103.7%	118.25%	153.54%	743.9%
W.o. p. risk	0.0016	145.2%	217.83%	450.77%	821.12%	5307.81%

Table 4.3: The time 0 ratio of std. of the hedging error with and without population basis risk. The ratios are computed in the same way as in Table 4.2.

static hedging strategy would significantly increase the hedger's hedging error. Realistic pension/annuity liabilities typically involve longer planning and payment horizons, as well as much more diversified pension members. As a result, the effect of the trading frequency is likely to be more profound in reality.

The effect of population basis risk

In this chapter we evaluate the performance of the hedging strategy with population basis risk. Now we look at the hedging performance of the strategies in the absence of population basis risk. In particular, we let both the reference population and the portfolio-specific population be the Australian males population, and report the results in Table 4.3. The results where both the reference population and the portfolio specific population is the Swedish males population are similar, and are thus omitted. We see that, in the absence of population basis risk, the standard deviation of the hedging error decreases drastically, especially for the benchmark case. The reason is that the hedging error resulted from the population basis risk is relatively invariant to the trading frequency of the longevity bond. Therefore, the existence of the population basis risk substantially lowers the ratios of the standard deviations. In situation without population basis risk, the hedging strategies become much more effective, and the effect of trading frequency becomes much more profound.

Sensitivity with respect to the longevity risk premium

As discussed above, the longevity-linked capital market is still at its infancy, and existing market price data is limited. At the current stage, a few methods to calibrate the longevity risk premium from existing market data are proposed in literature, such as the the Wang transform (Wang 2002; Lin and Cox 2005; Lin and Cox 2008), and the Sharpe Ratio method (Cairns et al. 2005; Bayraktar et al. 2009). However, as shown in Bauer et al. (2010), when calibrated to the UK annuity quote data, the above methods yield very different risk premiums. Therefore, it is important that the results produced by our model are robust to the choice of longevity risk premium values.

In order to test the sensitivity of our hedging strategy to the choice of longevity risk premiums, we evaluate the performance of the hedging strategies under two alternative sets of longevity risk premium values, $\tilde{\lambda}_{\mu,1} = \frac{1}{4}\lambda_\mu$ and $\tilde{\lambda}_{\mu,2} = 4\lambda_\mu$. The time 0 optimal standard error of the hedger's hedging error in the benchmark case, as well as the ratios under the other strategies are reported in Table 4.5. We see that, first, the value function and the ratios are robust to the change of the longevity risk premium. For example, the time 0 optimal standard error in the benchmark case changes by only 4.4% when the longevity risk premium becomes 16 times larger. Second, though only slightly, the value function of the benchmark case increases with the longevity risk premium. The intuition is that, as the longevity risk premium increases, the variability of the market value of the hedger's liabilities increases. More formally, as the longevity risk premium increases, the term $\frac{\partial G(s)}{\partial X_s}$ in the optimal strategy (4.38) decreases for all $s \in [t, T_0]$, and the time 0 variance of the hedger's net asset value, as given by (4.42), increases.[54] Moreover, the ratios of the standard deviations of other strategies change in the same direction as the longevity risk premium. Therefore, it seems that the time 0 standard deviation of the hedger's hedging error is more sensitive to the longevity risk premium for the constrained strategies. Finally, the time 0 standard deviation generated by the interest-only strategy is not affected by the change of the longevity risk premium, thus the ratio in the interest-only case decreases slightly as the longevity risk premium increases.

To sum up, our numerical results show that the effect of the trading frequency of

[54]Specifically, the explicit expression of $\frac{\partial G(s)}{\partial X_s}$ is given in Eqs (4.93) to (4.95). When the longevity risk premium increases, the term $\tilde{\alpha}_\mu(t, T_0, s)$ increases for all $0 \neq t \neq s \neq T$, while all other terms remain unchanged.

	Benchmark	1-year	2-year	5-year	Static	Interest-only
$\widetilde{\lambda}_{\mu,1}$	0.0159	101.07%	103.65%	118.04%	152.95%	744.395%
λ_μ	0.0160	101.09%	103.7%	118.25%	153.54%	743.9%
$\widetilde{\lambda}_{\mu,2}$	0.0166	101.18%	103.96%	119.30%	156.4%	743.57%

Table 4.4: The time 0 ratio of std. of the hedging error under different longevity risk premiums. In particular, $\widetilde{\lambda}_{\mu,1} = \frac{1}{4}\lambda_\mu$ and $\widetilde{\lambda}_{\mu,2} = 4\lambda_\mu$. The ratios are computed in the same way as in Table 4.2.

	Benchmark	1-year	2-year	5-year	Static	Interest-only
Discounted	0.0160	1.0109	1.0370	1.1825	1.5354	7.4390
Undiscounted	0.0192	1.0109	1.0371	1.1827	1.5360	7.4389

Table 4.5: The time 0 std. of the discounted and undiscounted hedging error in the benchmark case, and the ratio of the standard deviation in the other cases.

dynamic longevity risk hedging are robust to the change of the longevity risk premium.

The undiscounted hedging error

In this chapter we consider the hedger's discounted hedging error for the tractability of the optimization problem. Here, we calculate the standard deviations of the hedger's undiscounted hedging error under the optimal hedging strategies derived in this section. In other words, we compute

$$\mathrm{Var}_0[w^*(T_0) - Y(T_0)], \tag{4.67}$$

where $w^*(T_0)$ is the hedger's T_0 wealth under the optimal strategy.

In Table 4.5, we report the optimal standard deviation from the benchmark case, as well as the ratios of the standard deviation, for both the hedger's discounted and undiscounted hedging error. First, we see that the standard deviation in the undiscounted case is larger than the standard deviation in the discounted case. Second, more importantly, the ratios of the standard deviation are very similar in these two cases. Therefore, although the optimal strategies we derive is to minimize the standard deviation (variance) of the hedger's discounted hedging error, their relative performance remains almost the same when we do not discount.

4.7. Conclusion

In this chapter we study the dynamic hedging problem of a portfolio of liabilities exposed to longevity risk. In particular, we consider the case where a pension sponsor or an annuity provider (a hedger) wishes to minimize the variance of her hedging error, defined as the deviation of the market value of her investments in a money market account, zero coupon bonds, and longevity bonds, from the market value of her liabilities.

Closed-form optimal hedging strategies are obtained for the minimum-variance criterion under a forward mortality framework. In particular, time-consistent strategies are derived in the benchmark case and a liquidity constrained case. In the benchmark case, the hedger can rebalance both the zero coupon bonds and the longevity bonds continuously, while in the liquidity constrained case, the hedger can only rebalance the longevity bonds at a deterministic lower frequency.

The optimal hedges are evaluated in a numerical analysis with a Hull-White specification and parameter estimates from existing literature. We find that, compared with the benchmark case, limiting the trading of the longevity bond to a 2-year frequency only leads to a slight increase of the variance of the hedging error. Moreover, the dynamic hedging strategies, even when the hedger can only trade the longevity bond at a 5-year frequency, still significantly outperform the static hedging strategy (of the longevity bond).

There are several directions for future research. First, in this chapter we evaluate the hedging strategy using one parametrization of mortality process. To incorporate model risk, we would need to study forward mortality models with different parametric specifications, and evaluate the impact of different model specifications on the hedging effectiveness. Second, it would be interesting to extend the limitation of trading frequency to more realistic setups. For example, we may extend the deterministic trading times to stochastic trading times, or consider the case where the hedger faces a (stochastic) trading constraint regarding the amount of longevity-linked derivatives at each period. Third, in this chapter we evaluate the performance of the hedging strategies using instruments with the same maturities as the liabilities. However, at the currently stage, it might be difficult for the hedger to find counterparties to trade longevity-linked derivatives with a long maturity. Therefore, for a practical point of view, it would be interesting to

evaluate the impact of lacking hedging instruments with long maturities on the hedging of longevity risk. Moreover, in this case, it might be interesting to see what type of hedging instrument may yield the most satisfying hedging result. Fourth, the dynamic hedging portfolio could be interpreted as a "replicating portfolio" of the market-value of the longevity liabilities. The "replicating portfolio" can be used in applications such as calculating the solvency capital requirement.

4.8. Appendix

4.8.1. Proof of Proposition 1

For any $T \in \tilde{\mathcal{T}}$, we have, from (4.26),

$$dL(t,T) = E_t^Q[{}_Tp(x_0,0,rp)]dB(t,T) + B(t,T)dE_t^Q[{}_Tp(x_0,0,rp)]. \tag{4.68}$$

$dB(t,T)$ is given in Eq (4.24). Denote by $\mu(t,s,x_0,1)$, $\hat{\mu}(t,1)$, $a_\mu(u,t,1)$, and $\sigma_\mu(u,t,1)$ the first entry (row) of $\mu(t,s,x_0)$, $\hat{\mu}(t)$, $a_\mu(u,t)$, and $\sigma_\mu(u,t)$ for all $t,s \in [0,T]$. For the differential of $E_t^Q[{}_{T-t}p(x_0+t,t,rp)]$,

From (4.12), (4.13), and (4.21), we have

$$\begin{aligned}
&\log(E_t^Q[{}_{T-t}p(x_0+t,rp)]) \\
&= -\int_t^T [\mu(t,s,x_0,1) - \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)du]ds \\
&= -\int_t^T \mu(0,s,x_0,1)ds - \int_t^T \int_0^t a_\mu(u,s,1)duds - \int_t^T \int_0^t \sigma_\mu(u,s,1)dW_\mu^P(u)ds \\
&+ \int_t^T \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds.
\end{aligned} \tag{4.69}$$

(4.69) can be further written as

$$\begin{aligned}
&= -\int_0^T \mu(0,s,x_0,1)ds - \int_0^t \int_u^T a_\mu(u,s,1)dsdu - \int_0^t \int_u^T \sigma_\mu(u,s,1)dsdW_\mu^P(u) \\
&+ \int_0^t \mu(0,s,x_0,1)ds + \int_0^t \int_u^t a_\mu(u,s,1)dsdu + \int_0^t \int_u^t \sigma_\mu(u,s,1)dsdW_\mu^P(u) \\
&+ \int_t^T \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds,
\end{aligned} \tag{4.70}$$

where we interchange the integration order of $\int_t^T \int_0^t a_\mu(u,s,1)duds$ and $\int_t^T \int_0^t \sigma_\mu(u,s,1)dW_\mu^P(u)ds$, and decompose the inner-integration in two parts. Using (4.19), (4.20), and the fact that

$$\begin{aligned}\int_0^t \int_u^t \sigma_\mu(u,s,1)dsdW_\mu^P(u) &= \int_0^t \int_0^t \sigma_\mu(u,s,1)\mathbf{1}_{u\le s}dsdW_\mu^P(u)\\ &= \int_0^t \int_0^t \sigma_\mu(u,s,1)\mathbf{1}_{u\le s}dW_\mu^P(u)ds\\ &= \int_0^t \int_0^s \sigma_\mu(u,s,1)dW_\mu^P(u)ds, \qquad (4.71)\end{aligned}$$

we can further rewrite (4.70) as

$$\begin{aligned}&= \log(E_0^Q[{}_Tp(x_0,rp)]) - \int_0^T \int_0^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds + \int_t^T \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds\\ &- \int_0^t \int_u^T a_\mu(u,s,1)dsdu - \int_0^t \int_u^T \sigma_\mu(u,s,1)dsdW_\mu^P(u)\\ &+ \int_0^t \mu(0,s,x_0,1)ds + \int_0^t \int_0^s a_\mu(u,s,1)duds + \int_0^t \int_0^s \sigma_\mu(u,s,1)dW_\mu^P(u)ds\\ &= \log(E_0^Q[{}_Tp(x_0,rp)]) - \int_0^T \int_0^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds + \int_t^T \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds\\ &- \int_0^t \int_u^T a_\mu(u,s,1)dsdu - \int_0^t \int_u^T \sigma_\mu(u,s,1)dsdW_\mu^P(u) + \int_0^t \hat{\mu}(s,1)ds. \qquad (4.72)\end{aligned}$$

Denote by

$$\begin{aligned}\zeta_L(t,T) = &- \int_0^T \int_0^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds + \int_t^T \int_t^s \sigma_\mu(u,s,x_0,1)\lambda_\mu(u)duds\\ &- \int_0^t \int_u^T a_\mu(u,s,1)dsdu - \int_0^t \int_u^T \sigma_\mu(u,s,1)dsdW_\mu^P(u) + \int_0^t \hat{\mu}(s,1)ds, \qquad (4.73)\end{aligned}$$

then $E_t^Q[{}_{T-t}p(x_0+t,t,rp)]$ can be written as $E_0^Q[{}_Tp(x_0,0,rp)]\exp(\zeta_L(t,T))$. The differential of $\zeta_L(t,T)$ is given by

$$\begin{aligned}d\zeta_L(t,T) =&[\hat{\mu}(t,1) - \int_t^T \sigma_\mu(t,s,x_0,1)\lambda_\mu(t)ds - \int_t^T a_\mu(t,s,1)ds]dt\\ &-\{\int_t^T \sigma_\mu(t,s,x_0,1)ds\}dW_\mu^P(t). \qquad (4.74)\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
&dE_t^Q[{}_{T-t}p(x_0+t,t,rp)]\\
=&E_0^Q[{}_Tp(x_0,0,rp)]d\exp(\zeta_L(t,T))\\
=&E_0^Q[{}_Tp(x_0,0,rp)]\exp(\zeta_L(t,T))\{\hat{\mu}(t,1)-\int_t^T\sigma_\mu(t,s,x_0,1)\lambda_\mu(t)ds-\int_t^T a_\mu(t,s,1)ds\\
&+\frac{1}{2}(\int_t^T\sigma_\mu(t,s,x_0,1)ds)(\int_t^T\sigma_\mu(t,s,x_0,1)ds)'\}dt\\
&-E_0^Q[{}_Tp(x_0,0,rp)]\exp(\zeta_L(t,T))\{\int_t^T\sigma_\mu(t,s,x_0,1)ds\}dW_\mu^P(t)\\
=&E_t^Q[{}_{T-t}p(x_0+t,t,rp)]\{[\hat{\mu}(t,1)-\int_t^T\sigma_\mu(t,s,x_0,1)\lambda_\mu(t)ds]dt\\
&-[\int_t^T\sigma_\mu(t,s,x_0,1)ds]dW_\mu^P(t)\},
\end{aligned}
\tag{4.75}
$$

and the differential of $E_t^Q[{}_Tp(x_0,0,rp)]$ is given by

$$
\begin{aligned}
&dE_t^Q[{}_Tp(x_0,0,rp)]\\
=&d_tp(x_0,0,rp)E_t^Q[{}_{T-t}p(x_0+t,t,rp)]\\
=&de^{-\int_0^t\hat{\mu}(s,1)ds}E_t^Q[{}_{T-t}p(x_0+t,t,rp)]\\
=&-\hat{\mu}(t,1)e^{-\int_0^t\hat{\mu}(s,1)ds}E_t^Q[{}_{T-t}p(x_0+t,t,rp)]+e^{-\int_0^t\hat{\mu}(s,1)ds}dE_t^Q[{}_{T-t}p(x_0+t,t,rp)]\\
=&E_t^Q[{}_Tp(x_0,0,rp)]\{-[\int_t^T\sigma_\mu(t,s,x_0,1)\lambda_\mu(t)ds]dt-[\int_t^T\sigma_\mu(t,s,x_0,1)ds]dW_\mu^P(t)\}.
\end{aligned}
\tag{4.76}
$$

The differential of $L(t,T)$ can then be obtained by combining (4.76) and the dynamics of the zero coupon bond.

4.8.2. Proof of Theorem 2

In the presence of the liquidity constraint, $\hat{\boldsymbol{u}}_1$ is constant in the interval $[t_i,t_{i+1})$. In this case, we solve the optimization problem in two steps: we first solve the optimal $\hat{\boldsymbol{u}}_2$ as a function of $\hat{\boldsymbol{u}}_1$, then we solve the optimal $\hat{\boldsymbol{u}}_1$ recursively for each $[t_i,t_{i+1})$ (with $t_{n+1}=T_0$). For a fixed $\hat{\boldsymbol{u}}_1$, in terms of Wong et al. (2014), we have $Y(t)\equiv w(t)$, where $w(t)$ is given in (4.30) with $\hat{\boldsymbol{u}}_1(s)=\hat{\boldsymbol{u}}_1(t_i)$ for $s\in[t_i,t_{i+1})$, $i=1,2,\cdots,n$. Moreover, we

have[55]

$$\Gamma(t) \equiv G(\hat{\boldsymbol{u}}_1, t) = E_t[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \hat{\boldsymbol{u}}' a_w(s) ds - e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)],$$

where $G(\hat{\boldsymbol{u}}_1, t)$ emphasizes the dependence of G on $\hat{\boldsymbol{u}}_1$. Again, the risk aversion parameter ϕ in Wong et al. (2014) is set to be 0. In the first step, we only need to optimize with respect to $\hat{\boldsymbol{u}}_2$, which is continuously rebalanced. Repeating Proposition 3.1 and 3.2 in Wong et al. (2014), we obtain the optimal $\hat{\boldsymbol{u}}_2$:

$$\begin{aligned}\hat{\boldsymbol{u}}_2(t) = &- (\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_X(t)' \frac{\partial G(\hat{\boldsymbol{u}}_1, t)}{\partial X(t)} \\ &-(\sigma_w(t,2)\sigma_w(t,2)')^{-1}\sigma_w(t,2)\sigma_w(t,1)'\hat{\boldsymbol{u}}_1(t_i), \quad t \in [t_i, t_{i+1}),\end{aligned} \tag{4.77}$$

with $G(\hat{\boldsymbol{u}}_1, t)$ represented as

$$\begin{aligned}G(\hat{\boldsymbol{u}}_1, t) = &\sum_{j=i+1}^{n} E_t^{\bar{Q}}[\int_{t_j}^{t_{j+1}} e^{-\int_t^s r(\tau)d\tau} \hat{\boldsymbol{u}}_1(t_j) a_w(s,1) ds] - E_t^{\bar{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau} Y_{T_0}] \\ &+\hat{\boldsymbol{u}}_1(t_i)' E_t^{\bar{Q}}[\int_t^{t_{i+1}} e^{-\int_t^s r(\tau)d\tau} a_w(s,1) ds] \\ \equiv &G_i(\hat{\boldsymbol{u}}_1, t) + \hat{\boldsymbol{u}}_1(t_i)' H_i(t),\end{aligned} \tag{4.78}$$

Next, we proceed to the second step and solve the optimal $\hat{\boldsymbol{u}}_1$. For each $i = 1, \cdots, n$, the value function at t_i is given by

$$J_{t_i} = \min_{\hat{\boldsymbol{u}}_1(t_i)} \text{Var}_{t_i}\{e^{-\int_{t_i}^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))\}, \tag{4.79}$$

where $w^*(T_0)$ is the wealth at T_0 under $\hat{\boldsymbol{u}}^*$ from t_{i+1} to T_0. To derive the optimal $\hat{\boldsymbol{u}}_1^*$, we make use of an alternative formulation of the law of total variance proposed in Proposition 2 from Basak and Chabakauri (2012) (Equations A15 – A16). From the law of total variance in (4.35), with an infinitesimally small time interval ϵ, we obtain the following differential form

$$\begin{aligned}0 = &E_t[d\ \text{Var}_s\{e^{-\int_s^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))\} \\ &+\text{Var}_s\{d\ e^{-\int_t^s r(\tau)d\tau} E_s[e^{-\int_s^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))]\}],\end{aligned} \tag{4.80}$$

[55] Again, for notational simplicity, we subtract the subscript $X(t)$ from the relevant quantities.

Letting $t = t_i$, and integrating (4.80) from t_i to t_{i+1}, we have (following the notations in Basak and Chabakauri (2012))

$$\begin{aligned}
&\mathrm{Var}_{t_i}\{e^{-\int_{t_i}^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))\} \\
=&E_{t_i}[\mathrm{Var}_{t_{i+1}}\{e^{-\int_{t_{i+1}}^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))\}] \\
+&E_{t_i}[\int_{t_i}^{t_{i+1}} \frac{\mathrm{Var}_s\{d\ e^{-\int_{t_i}^{s} r(\tau)d\tau} E_s[e^{-\int_s^{T_0} r(\tau)d\tau}(w(T_0) - Y(T_0))]\}}{ds} ds]. \qquad (4.81)
\end{aligned}$$

For $t \in [t_i, t_{i+1})$, let $w^*(T_0)$ be the wealth at T_0 under $\hat{\boldsymbol{u}}^*$ from t_{i+1} to T_0. Denote by $\tilde{\boldsymbol{u}}_1$ the strategy where $\tilde{\boldsymbol{u}}_1 = \hat{\boldsymbol{u}}_1^*$ for $t \in [t_{i+1}, T_0]$ and $\tilde{\boldsymbol{u}}_1 = \hat{\boldsymbol{u}}_1$ for a fixed $\hat{\boldsymbol{u}}_1$ for $t \in [t_i, t_{i+1})$, and denote by $\tilde{\boldsymbol{u}}_2(\hat{\boldsymbol{u}}_1)$ the strategy where $\tilde{\boldsymbol{u}}_2(\hat{\boldsymbol{u}}_1) = \hat{\boldsymbol{u}}_2^*$ for $t \in [t_{i+1}, T_0]$ and takes the form (4.77) for the given $\hat{\boldsymbol{u}}_1$ for $t \in [t_i, t_{i+1})$. Moreover, denote by $\tilde{\boldsymbol{u}}(s) = (\tilde{\boldsymbol{u}}_1(s)', \tilde{\boldsymbol{u}}_2(\hat{\boldsymbol{u}}_1, s)')'$. From the integral form of $w^*(T_0)$,

$$\begin{aligned}
w^*(T_0)e^{-\int_t^{T_0} r(\tau)d\tau} =& w(t) + \int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \tilde{\boldsymbol{u}}(s)' a_w(s) ds \\
&+ \int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \tilde{\boldsymbol{u}}(s)' \sigma_w(s) dW(s), \qquad (4.82)
\end{aligned}$$

we have

$$E_t[e^{-\int_t^{T_0} r(\tau)d\tau} w^*(T_0)] = w(t) + E_t[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \tilde{\boldsymbol{u}}(s)' a_w(s) ds].$$

Therefore, applying Ito's lemma, we have

$$\begin{aligned}
&d\ e^{-\int_{t_i}^t r(\tau)d\tau} E_t[e^{-\int_t^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))] \\
=&\{...\}dt + e^{-\int_{t_i}^t r(\tau)d\tau}[\sigma_X(t)\frac{\partial \tilde{E}_i(\hat{\boldsymbol{u}}_1, t)}{\partial X(t)} + \tilde{\boldsymbol{u}}(s)'\sigma_w(t)]dW(t), \qquad (4.83)
\end{aligned}$$

where

$$\tilde{E}_i(\hat{\boldsymbol{u}}_1, t) \equiv E_t[\int_t^{T_0} e^{-\int_t^s r(\tau)d\tau} \tilde{\boldsymbol{u}}(s)' a_w(s) ds - e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)]$$

for $t \in [t_i, t_{i+1})$. The term $\tilde{E}_i(\hat{\boldsymbol{u}}_1, t)$ is equal to $G(t)$. To see this, apply the Feynman-Kac Theorem to $\tilde{E}_i(\hat{\boldsymbol{u}}_1, t)$, and substitute $\tilde{\boldsymbol{u}}_2(\hat{\boldsymbol{u}}_1, s)$. Applying again the Feynman-Kac Theorem to the resulting partial differential equation, and recognizing the terminal condition

$\tilde{E}_i(\hat{\boldsymbol{u}}_1, t_{i+1}) = G_i(\hat{\boldsymbol{u}}_1, t_{i+1})$, we have

$$\begin{aligned}\tilde{E}_i(\hat{\boldsymbol{u}}_1, t) =& E_t^{\bar{Q}}[G_i(\hat{\boldsymbol{u}}_1, t_{i+1}) + \int_t^{t_{i+1}} e^{-\int_t^s r(\tau)d\tau} \hat{\boldsymbol{u}}_1(t_i)' a_w(s,1)ds] \\ =& G_i(\hat{\boldsymbol{u}}_1, t) + \hat{\boldsymbol{u}}_1(t_i)' E_t^{\bar{Q}}[\int_t^{t_{i+1}} e^{-\int_t^s r(\tau)d\tau} a_w(s,1)ds] \\ =& G(\hat{\boldsymbol{u}}_1, t). \end{aligned} \tag{4.84}$$

Substituting (4.83) into (4.81) and computing $\text{Var}_s\{d\, e^{-\int_{t_i}^s r(\tau)d\tau} E_s[e^{-\int_s^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))]$ yields

$$\begin{aligned}&\text{Var}_{t_i}\{e^{-\int_{t_i}^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))\} \\ =& E_{t_i}[\text{Var}_{t_{i+1}}\{e^{-\int_{t_{i+1}}^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))\}] \\ +& E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^s r(\tau)d\tau}[\sigma_X(s)\frac{\partial G(\hat{\boldsymbol{u}}_1, s)}{\partial X(s)} + \tilde{\boldsymbol{u}}(s)'\sigma_w(s)] \\ & [\sigma_X(s)\frac{\partial G(\hat{\boldsymbol{u}}_1, s)}{\partial X(s)} + \tilde{\boldsymbol{u}}(s)'\sigma_w(s)]' ds] \end{aligned} \tag{4.85}$$

Substituting the integral form of $w^*(T_0)$ into the time t_{i+1} value function, $\text{Var}_{t_{i+1}}$ $\{e^{-\int_{t_{i+1}}^{T_0} r(\tau)d\tau}(w^*(T_0) - Y(T_0))\}$, we see that it does not depend on $w(t_{i+1})$, and thus does not depend on the strategy $\hat{\boldsymbol{u}}_1$ on $[t_i, t_{i+1})$. Therefore, we only need to focus on the second part on the right hand side of (4.85). Substituting the optimal $\hat{\boldsymbol{u}}_2^*$ into (4.85), and taking the derivative with respect to $\hat{\boldsymbol{u}}_1$, we obtain

$$\hat{\boldsymbol{u}}_1^*(t_i) = (E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^s r(\tau)d\tau} A_i(s)ds])^{-1} E_{t_i}[\int_{t_i}^{t_{i+1}} e^{-2\int_{t_i}^s r(\tau)d\tau} B_i(s)ds]. \tag{4.86}$$

where the $A_i(s)$ and $B_i(s)$ matrices are given in (4.51).

4.8.3. Optimal hedging strategies under the Hull-White specification

In this section we derive the explicit form of the benchmark and constrained minimum-variance optimal hedging strategy under the Hull-White specification given in Section 4.6. First, we proceed to solve $\frac{\partial G(t)}{\partial X(t)}$. Recall that $G(t) = E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau} Y(T_0)]$ and

$$Y(T_0) =_{T_0} p(x_0, 0, pp) \int_{T_0}^{T_1} B(T_0, s) E_t^Q[_{s-T_0}p(x_0 + T_0, T_0, pp)]ds.$$

Wee can see that $Y(T_0)$ only depends on $W^Q(s)$ for $s \in [T_0, T_1]$, while, on the other hand, $e^{-\int_t^{T_0} r(\tau)d\tau}$ only depend on $W^Q(s)$ for $t \in [t, T_0]$. Therefore, by the independence of $W_r^Q(t)$ and $W_\mu^Q(t)$, $G(t)$ can be written as

$$\begin{aligned}
&- E_t^{\tilde{Q}}[e^{-\int_0^{T_0}\hat{\mu}(\nu,x_0,pp)d\nu}]E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}]\int_{T_0}^{T_1} E_t^{\tilde{Q}}[B(T_0,s)]E_t^{\tilde{Q}}\{E_{T_0}^Q[{}_{s-t}p(x_0+t,t,pp)]\}ds \\
&= -{}_{tp(x_0,0,pp)}\, E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}]\int_{T_0}^{T_1} E_t^{\tilde{Q}}[B(T_0,s)] \\
&\quad E_t^{\tilde{Q}}\{{}_{T_0-t}p(x_0+t,t,pp)E_{T_0}^Q[{}_{s-T_0}p(x_0+T_0,T_0,pp)]\}ds. \qquad (4.87)
\end{aligned}$$

The second equality of (4.87) holds since ${}_{T_0}p(x_0,0,pp) = {}_tp(x_0,0,pp){}_{T_0-t}p(x_0+t,t,pp)$, and ${}_{T_0}p(x_0+t,t,pp)$ and $E_{T_0}^Q[{}_{s-T_0}p(x_0+T_0,T_0,2)$ are independent. Before giving the explicit solution of $\frac{\partial G(t)}{\partial X(t)}$, we first derive the dynamics of $X(t,x_0)$ under equivalent martingale measure Q.

Proposition 4. *Under the specification of the volatilities given in* (4.63) *and* (4.64)*, the state variable $X(t,x_0)$ is 7 dimensional. In particular, it consists of $r(t)$, $\hat{\mu}(t,x_0)$, and $\eta_{kj}(t,x_0)$ with $k = rp, pp$, $j = 2,3$. The dynamics of these state variables under Q are*

$$\begin{aligned}
dr(t) =&[\theta_r^Q(t) - \kappa r(t)]dt + \beta dW_r^Q(t) \\
d\hat{\mu}(t,x_0,k) =&[\theta_{k,\mu}^Q(t) + \sum_{j=2}^{3}(\omega_{k1}-\omega_{kj})\eta_{kj}(t) - \omega_{k1}\hat{\mu}(t,x_0,k)]dt + \sum_{j=1}^{3} c_{kj}(x_0)dW_{j,\mu}^Q(t) \\
d\eta_{kj}(t) =&[c_{kj}\lambda_{j,\mu}(t) - \omega_{kj}\eta_{kj}(t)]dt + c_{kj}dW_{j,\mu}^Q(t), \qquad (4.88)
\end{aligned}$$

where

$$\begin{aligned}
\theta_r^Q(t) =&\kappa f(0,t) + f_2(0,t) + \frac{\beta^2}{2\kappa}(1-e^{-2\kappa t}) \\
\theta_{k,\mu}^Q(t) =&\mu_2(0,t,k) + \omega_{k1}\mu(0,t,pp) + \frac{c_{k1}^2}{2\omega_{k1}}(1-e^{-2\omega_{k1}t}) \\
&+ \sum_{j=2}^{3}[\frac{c_{kj}^2}{\omega_{kj}}e^{-2\omega_{kj}t}(1-e^{-\omega_{kj}t}) + \frac{\omega_{k1}c_{kj}^2}{2\omega_{kj}^2}(1-e^{-\omega_{kj}t})^2 + c_{kj}\lambda_{j,\mu}(t)] \qquad (4.89)
\end{aligned}$$

for $k = rp, pp$ and $j = 2,3$. In (4.89)*, $\mu_2(0,t,k)$ is the derivative of $\mu(0,t,k)$ with respective to t.*

Proof. See Appendix 4.8.3. □

After having the dynamics of $X(t, x_0)$ under the measure Q, we are able to derive explicit representation of $G(t)$, as well as the derivative $\frac{\partial G(t)}{\partial X(t)}$. The results are given in the next Proposition.

Proposition 5. *The conditional expectations in $G(t)$ can be solved as*

$$E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}] = e^{\alpha_r(t,T_0)+\beta_r(t,T_0)r(t)}, \tag{4.90}$$

$$E_t^{\tilde{Q}}[B(T_0, s)] = e^{\tilde{\alpha}_r(t,T_0,s)+\beta_r(T_0,s)e^{-\kappa(T_0-t)}r(t)}, \tag{4.91}$$

and

$$\begin{aligned} &E_t^{\tilde{Q}}\{{}_{T_0}p(x_0+t,t,pp)E_{T_0}^{Q}[{}_{s-T_0}p(x_0+T_0,T_0,pp)]\} \\ =&e^{\tilde{\alpha}_\mu(t,T_0,s)+\tilde{\beta}_\mu(t,T_0,s)\hat{\mu}(t,2)+\tilde{\beta}_{2,\eta}(t,T_0,s)\eta_{22}(t)+\tilde{\beta}_{3,\eta}(t,T_0,s)\eta_{23}(t)}. \end{aligned} \tag{4.92}$$

Proof. For the proof and the representation of the relevant quantities, see Appendix 4.8.3. □

With the help of Proposition 5, $\frac{\partial G(t)}{\partial X(t)}$ is given by

$$\begin{aligned} \frac{\partial G(t)}{\partial r(t)} =& -{}_tp(x_0,0,pp)e^{\alpha_r(t,T_0)+\beta_r(t,T_0)r(t)}\int_{T_0}^{\tilde{T}}(\beta_r(t,T_0)+\beta_r(T_0,s)e^{-\kappa(T_0-t)}) \\ &e^{\tilde{\alpha}_r(t,T_0,s)+\beta_r(T_0,s)e^{-\kappa(T_0-t)}r(t)}e^{\tilde{\alpha}_\mu(t,T_0,s)+\tilde{\beta}_\mu(t,T_0,s)\hat{\mu}(t,2)+\tilde{\beta}_{2,\eta}(t,T_0,s)\eta_{22}(t)\tilde{\beta}_{3,\eta}(t,T_0,s)\eta_{23}(t)}ds, \end{aligned} \tag{4.93}$$

$$\begin{aligned} \frac{\partial G(t)}{\partial \hat{\mu}(t,pp)} =& -{}_tp(x_0,0,pp)e^{\alpha_r(t,T_0)+\beta_r(t,T_0)r(t)}\int_{T_0}^{\tilde{T}}\tilde{\beta}_\mu(t,T_0,s)e^{\tilde{\alpha}_r(t,T_0,s)+\beta_r(T_0,s)e^{-\kappa(T_0-t)}r(t)} \\ &e^{\tilde{\alpha}_\mu(t,T_0,s)+\tilde{\beta}_\mu(t,T_0,s)\hat{\mu}(t,2)+\tilde{\beta}_{2,\eta}(t,T_0,s)\eta_{22}(t)+\tilde{\beta}_{3,\eta}(t,T_0,s)\eta_{23}(t)}ds, \end{aligned} \tag{4.94}$$

and

$$\begin{aligned} \frac{\partial G(t)}{\partial \eta_{2j}(t)} =& -{}_tp(x_0,0,pp)e^{\alpha_r(t,T_0)+\beta_r(t,T_0)r(t)}\int_{T_0}^{\tilde{T}}\tilde{\beta}_{j,\eta}(t,T_0,s)e^{\tilde{\alpha}_r(t,T_0,s)+\beta_r(T_0,s)e^{-\kappa(T_0-t)}r(t)} \\ &e^{\tilde{\alpha}_\mu(t,T_0,s)+\tilde{\beta}_\mu(t,T_0,s)\hat{\mu}(t,2)+\tilde{\beta}_{2,\eta}(t,T_0,s)\eta_{22}(t)+\tilde{\beta}_{3,\eta}(t,T_0,s)\eta_{23}(t)}ds, \end{aligned} \tag{4.95}$$

for $j = 2, 3$.

For the constrained $\hat{\boldsymbol{u}}$, we need to compute the conditional expectations $E_{t_i}[\frac{\partial G(s)}{\partial X(s)}]$ for $s \in [t_i, t_{i+1}]$. This can be done in a similar way as in Proposition 5.

Proof of Proposition 4

In this section, we omit the subscript x_0 in the mortality process and related parameters for the simplicity of notation. However, we still recognize that all mortality processes in this section are contingent on the cohort x_0. From (4.3), (4.8), and (4.19), we can write $r(t)$ as

$$\begin{aligned} r(t) =& f(0,t) + \frac{\beta^2}{2\kappa^2}(e^{-\kappa t} - 1)^2 - \int_0^t \beta e^{-\kappa(t-s)} \lambda_r(s) ds + \int_0^t \beta e^{-\kappa(t-s)} \lambda_r(s) ds \\ &+ \beta e^{-\kappa t} \int_0^t e^{\kappa u} dW_r^{\tilde{Q}}(u) \\ =& f(0,t) + \frac{\beta^2}{2\kappa^2}(e^{-\kappa t} - 1)^2 + \beta e^{-\kappa t} \int_0^t e^{\kappa u} dW_r^{Q}(u). \end{aligned} \tag{4.96}$$

Denote by $g_r(t) = f(0,t) + \frac{\beta^2}{2\kappa^2}(e^{-\kappa t} - 1)^2$ and $R_r(t) = \int_0^t e^{\kappa u} dW_r^Q(u)$. We can write $r(t)$ as

$$r(t) = G_r(t, R_r) = g(t) + \beta e^{-\kappa t} R_r(t), \tag{4.97}$$

with $\frac{\partial G_r(t,R_r)}{\partial t} = g'(t) - \kappa\beta e^{-\kappa t} R_r(t)$, $\frac{\partial G_r(t,R_r)}{\partial R_r} = \beta e^{-\kappa t}$, $\frac{\partial^2 G_r(t,R_r)}{\partial R_r^2} = 0$. Together with the fact that $dR_r(t) = e^{\kappa t} dW_r^Q(t)$, we can write

$$dr(t) = [f_2(0,t) + \frac{\beta^2}{\kappa} e^{-\kappa t}(1 - e^{-\kappa t}) - \kappa\beta e^{-\kappa t} R_r(t)]dt + \beta e^{-\kappa t} e^{\kappa t} dW_r^Q(t). \tag{4.98}$$

Using the fact that $\kappa\beta e^{-\kappa t} R_r(t) = \kappa(r(t) - g_r(t))$, we can further write (4.98) as

$$dr(t) = [\theta_r^Q(t) - \kappa r(t)]dt + \beta dW_r^Q(t), \tag{4.99}$$

where

$$\theta_r^Q(t) = \kappa f(0,t) + f_2(0,t) + \frac{\beta^2}{2\kappa}(1 - e^{-2\kappa t}).$$

For the spot force of mortality process, we have

$$\begin{aligned}\hat{\mu}(t,k) =&\mu(0,t,k) + \frac{1}{2}\sum_{j=1}^{3}\frac{c_{kj}^2}{\omega_{kj}^2}(1-e^{-\omega_{kj}t})^2 - \sum_{j=1}^{3} c_{kj}e^{-\omega_{kj}t}\int_0^t e^{\omega_{kj}s}\lambda_{j,\mu}(s)ds \\ &+\sum_{j=1}^{3} c_{kj}e^{-\omega_{kj}t}\int_0^t e^{\omega_{kj}s}dW_{j,\mu}^{Q}(s), \end{aligned} \tag{4.100}$$

and

$$\eta_{kj}(t) = -\int_0^t c_{kj}e^{-\omega_{kj}(t-s)}\lambda_{j,\mu}(s)ds + \int_0^t c_{kj}e^{-\omega_{kj}(t-s)}dW_{j,\mu}^{Q}(s) \tag{4.101}$$

for $k = 1, 2$ and $j = 2, 3$. From (4.101) we directly have

$$d\eta_{kj}(t) = [-c_{kj}\lambda j, \mu(t) - \omega_{kj}\eta_{kj}(t)]dt + c_{kj}dW_{j,\mu}^{Q}(t). \tag{4.102}$$

Moreover, similar to the deviation of $dr(t)$, we have

$$d\hat{\mu}(t,k) = [\theta_{k,\mu}^{Q}(t) + \sum_{j=2}^{3}(\omega_{k1}-\omega_{kj})\eta_{kj}(t) - \omega_{k1}\hat{\mu}(t,k)]dt + \sum_{j=1}^{3} c_{kj}dW_{j,\mu}^{Q}(t), \tag{4.103}$$

with

$$\begin{aligned}\theta_{k,\mu}^{Q}(t) =&\mu_2(0,t,k) + \omega_{k1}\mu(0,t,2) - c_{k1}\lambda_{1,\mu}(t) + \frac{c_{k1}^2}{2\omega_{k1}}(1-e^{-2\omega_{k1}t}) \\ &+\sum_{j=2}^{3}[\frac{c_{kj}^2}{\omega_{kj}}e^{-\omega_{kj}t}(1-e^{-\omega_{kj}t}) + \frac{\omega_{k1}c_{kj}^2}{2\omega_{kj}^2}(1-e^{-\omega_{kj}t})^2 - c_{kj}\lambda_{j,\mu}(t)]. \end{aligned} \tag{4.104}$$

Proof of Proposition 5

For $E_t^{\tilde{Q}}[e^{-\int_t^{T_0} r(\tau)d\tau}]$, we can directly apply Eqs (2.3) to (2.6) and Appendix B in Duffie et al. (2000). In particular, using the notations in Duffie et al. (2000), we have $K_0(t) = \theta_r(t)$, $K_1 = -\kappa$, $H_0 = \beta^2$, $H_1 = 0$, $\rho_0 = 0$, and $\rho_1 = 1$. As a result, we have

$$\begin{aligned}\beta_r(t,T_0) =& -\frac{1}{\kappa}(1-e^{-\kappa(T_0-t)}) \\ \alpha_r(t,T_0) =& \int_t^{T_0}\theta_r(s)\beta_r(s,T_0)ds - \frac{\beta^2}{4\kappa}\beta_r^2(t,T_0) + \frac{\beta^2}{2\kappa^2}(T-t+\beta_r(t,T_0)) \end{aligned} \tag{4.105}$$

in Eq (4.90).

Using (4.90), we can write the second term as $E_t^{\tilde{Q}}[B(T_0,s)] = E_t^{\tilde{Q}}[e^{\alpha_r(T_0,s)+\beta_r(T_0,s)r(T_0)}]$.

Thus, we need to determine the distribution of $r(t)$ under $\tilde{Q}$. $r(T_0)$ can be written as

$$\begin{aligned} r(T_0) =& f(0,T_0) + \frac{\beta^2}{2\kappa^2}(e^{-\kappa T_0}-1)^2 + \beta e^{-\kappa T_0}\int_0^{T_0} e^{\kappa u} dW_r^{\tilde{Q}}(u) \\ =& e^{-\kappa(T_0-t)} r(t) + g_r(T_0) - e^{-\kappa(T_0-t)} g_r(t) + \beta e^{-\kappa T_0}\int_t^{T_0} e^{\kappa u} dW_r^{\tilde{Q}}(u), \end{aligned} \tag{4.106}$$

where

$$g_r(t) = f(0,t) + \frac{\beta^2}{2\kappa^2}(e^{-\kappa t}-1)^2. \tag{4.107}$$

From (4.106), we see that the distribution of $r(T_0)$ conditional on $r(t)$ under $\tilde{Q}$ is normal with mean and variance given as

$$\begin{aligned} E_t^{\tilde{Q}}[r(T_0)] =& e^{-\kappa(T_0-t)} r(t) + g_r(T_0) - e^{-\kappa(T_0-t)} g_r(t), \\ \mathrm{Var}_t^{\tilde{Q}}[r(T_0)] =& \frac{\beta^2}{2\kappa} e^{-2\kappa T_0}(e^{2\kappa T_0} - e^{2\kappa t}). \end{aligned} \tag{4.108}$$

Therefore, we have the representation of $E_t^{\tilde{Q}}[B(T_0,s)]$ in (4.91) with

$$\tilde{\alpha}_r(t,T_0,s) = \alpha_r(t,T_0) + \beta_r(T_0,s)(g_r(T_0) - e^{-\kappa(T_0-t)} g_r(t)) + \frac{\beta^2}{4\kappa}\beta_r^2(T_0,s)(1-e^{-2\kappa(T_0-t)}). \tag{4.109}$$

For the last term, $E_t^{\tilde{Q}}\{{}_{T_0}p(x_0+t,t,pp) E_{T_0}^{Q}[{}_{s-T_0}p(x_0+T_0,T_0,pp)]\}$, we first deal with two terms inside $E_t^{\tilde{Q}}\{.\}$, respectively. First, using the dynamics of $\hat{\mu}(t)$ under Q derived in Proposition 4, we can write $E_{T_0}^{Q}[{}_{s-T_0}p(x_0+T_0,T_0,pp)]$ as

$$\begin{aligned} & E_{T_0}^{Q}[{}_{s-T_0}p(x_0+T_0,T_0,pp)] \\ =& E_{T_0}^{Q}[e^{-\int_{T_0}^{s}\hat{\mu}(\tau,2)d\tau}] \\ =& e^{\alpha_\mu(T_0,s)+\beta_\mu(T_0,s)\hat{\mu}(T_0,2)+\beta_{2,\eta}(T_0,s)\eta_{22}(T_0)+\beta_{3,\eta}(T_0,s)\eta_{23}(T_0)}, \end{aligned} \tag{4.110}$$

with

$$\begin{aligned} \beta_\mu(T_0,s) =& \frac{1}{\omega_{21}}(e^{-\omega_{21}(s-T_0)}-1) \\ \beta_{j,\eta}(T_0,s) =& \frac{1}{\omega_{21}}(e^{-\omega_{2j}(s-T_0)} - e^{-\omega_{21}(s-T_0)}) + \frac{\omega_{2j}-\omega_{21}}{\omega_{21}\omega_{2j}}(1-e^{-\omega_{2j}(s-T_0)}) \\ \alpha_\mu(T_0,s) =& \int_{T_0}^{s}[\mathbf{K}_0(u)\beta(u,s) + \frac{1}{2}\beta(u,s)'\mathcal{H}_0(u)\beta(u,s)]du. \end{aligned} \tag{4.111}$$

The terms $\beta(u,s)$, $\mathbf{K}_0(u)$, and $\mathcal{H}_0(u)$ in (4.111) are

$$\beta(u,s) = (\beta_\mu(u,s), \beta_{2,\eta}(u,s), \beta_{3,\eta}(u,s))', \tag{4.112}$$

$$\mathbf{K}_0(u) = (\theta^Q_{2,\mu}(u), -c_{22}\lambda_{2,\mu}(u), -c_{23}\lambda_{3,\mu}(u))', \tag{4.113}$$

and

$$\mathcal{H}_0(u) = \begin{pmatrix} c_{21}^2 + c_{22}^2 + c_{23}^2 & c_{22}^2 & c_{23}^2 \\ c_{22}^2 & c_{22}^2 & 0 \\ c_{23}^2 & 0 & c_{23}^2 \end{pmatrix}, \tag{4.114}$$

respectively. $E_t^{\tilde{Q}}[{}_{T_0}p(x_0+t,t,pp)]$ can be solved similarly as $E^Q_{T_0}[{}_{s-T_0}p(x_0+T_0,T_0,2)]$. Denote by

$$\begin{aligned} \tilde{\lambda}_\mu \equiv & (\tilde{\lambda}_{1,\mu}(t), \tilde{\lambda}_{2,\mu}(t), \tilde{\lambda}_{3,\mu}(t))' \\ \equiv & \frac{\int_t^T \sigma_{\mu,1}(t,s)ds \int_t^T \sigma_{\mu,1}(t,s)ds' \lambda_\mu(t)}{\int_t^T \sigma_{\mu,1}(t,s)ds' \int_t^T \sigma_{\mu,1}(t,s)ds}, \end{aligned} \tag{4.115}$$

the dynamics of $\hat{\mu}(t,2)$, $\eta_{22}(t)$, and $\eta_{23}(t)$ are given by similar expressions as (4.102) to (4.104), with λ_μ replaced by $\tilde{\lambda}_\mu$. As a result, we have

$$\begin{aligned} & E_t^{\tilde{Q}}[{}_{T_0}p(x_0+t,t,pp)] \\ = & e^{\bar{\alpha}_\mu(t,T_0)+\bar{\beta}_\mu(t,T_0)\hat{\mu}(t,2)+\bar{\beta}_{2,\eta}(t,T_0)\eta_{22}(t)+\bar{\beta}_{3,\eta}(t,T_0)\eta_{23}(t)}, \end{aligned} \tag{4.116}$$

with $\bar{\beta}_\mu(t,s) = \beta_\mu(t,s)$ and $\bar{\beta}_{j,\eta}(t,s) = \beta_{j,\eta}(t,s)$ for all $0 \le t \le s \le T$ and $j = 2,3$. $\bar{\alpha}_\mu(t,T_0)$ has the same expression as $\alpha_\mu(t,T_0)$ in (4.111) with all λ_μ replaced by $\tilde{\lambda}_\mu$.

To explicitly solve $E_t^{\tilde{Q}}\{E^Q_{T_0}[{}_{s-T_0}p(x_0+T_0,2)]\}$, we need to know the time t conditional distribution of $\hat{\mu}(T_0,2)$ and $\eta_{2j}(T_0)$-s under $\tilde{Q}$. Denote by

$$\begin{aligned} g^{\tilde{Q}}_{2,\mu}(t) = & \mu(0,t,2) + \frac{1}{2}\sum_{j=1}^3 \frac{c_{2j}^2}{\omega_{2j}^2}(1-e^{-\omega_{2j}t})^2 - \sum_{j=1}^3 c_{2j}e^{-\omega_{2j}t}\int_0^t e^{\omega_{2j}s}\tilde{\lambda}_{j,\mu}(s)ds, \\ g^{\tilde{Q}}_{j,\eta}(t) = & -c_{2j}e^{-\omega_{2j}t}\int_0^t e^{\omega_{2j}s}\tilde{\lambda}_{j,\mu}(s)ds, \end{aligned} \tag{4.117}$$

we can write

$$\begin{aligned}\hat{\mu}(t,2) =& g^{\tilde{Q}}_{2,\mu}(t) + \sum_{j=1}^{3} c_{2j} e^{-\omega_{2j} t} \int_0^t e^{\omega_{2j} s} dW^{\tilde{Q}}_{j,\mu}(s) \\ \eta_{2,j}(t) =& g^{\tilde{Q}}_{j,\eta}(t) + c_{2j} e^{-\omega_{2j} t} \int_0^t e^{\omega_{2j} s} dW^{\tilde{Q}}_{j,\mu}(s). \end{aligned} \tag{4.118}$$

Therefore, given $\hat{\mu}(t)$, $\hat{\mu}(T_0, 2)$ and $\eta_{2j}(T_0)$-s can be written as

$$\begin{aligned}\eta_{2,j}(T_0) =& e^{-\omega 2j(T_0-t)} \eta_{2,j}(t) + g^{\tilde{Q}}_{j,\eta}(T_0) - e^{-\omega 2j(T_0-t)} g^{\tilde{Q}}_{j,\eta}(t) + c_{2j} e^{-\omega_{2j} T_0} \int_t^{T_0} e^{\omega_{2j} s} dW^{\tilde{Q}}_{j,\mu}(s) \\ \hat{\mu}(T_0,2) =& e^{-\omega 21(T_0-t)} \hat{\mu}(t,2) + g^{\tilde{Q}}_{2,\mu}(T_0) - e^{-\omega 21(T_0-t)} g^{\tilde{Q}}_{2,\mu}(t) \\ &+ \sum_{j=2}^{3} (e^{-\omega_{2j}(T_0-t)} - e^{-\omega_{21}(T_0-t)})(\eta_{2j}(t) - g^{\tilde{Q}}_{j,\eta}(t)) + \sum_{j=1}^{3} c_{2j} e^{-\omega_{2j} T_0} \int_t^{T_0} e^{\omega_{2j} s} dW^{\tilde{Q}}_{j,\mu}(s). \end{aligned} \tag{4.119}$$

Therefore, we see that $(\hat{\mu}(T_0, 2), \eta_{22}(T_0), \eta_{23}(T_0))$ is a tri-variate normal distributed variable with variance and covariance given by

$$\begin{aligned}\mathrm{Var}^{\tilde{Q}}_t[\eta_{2,j}(T_0)] =& \frac{c^2_{2j}}{2\omega_{2j}} [1 - e^{-2\omega_{2j}(T_0-t)}] \\ \mathrm{Var}^{\tilde{Q}}_t[\hat{\mu}(T_0,2)] =& \sum_{j=1}^{3} \frac{c^2_{2j}}{2\omega_{2j}} [1 - e^{-2\omega_{2j}(T_0-t)}] \\ \mathrm{Cov}^{\tilde{Q}}_t[\hat{\mu}(T_0,2), \eta_{2,j}(t)] =& \frac{c^2_{2j}}{2\omega_{2j}} [1 - e^{-2\omega_{2j}(T_0-t)}]. \end{aligned} \tag{4.120}$$

After solving both expectations, $E^{\tilde{Q}}_t\{{}_{T_0}p(x_0 + t, t, pp) E^{Q}_{T_0}[{}_{s-T_0}p(x_0 + T_0, T_0, pp)]\}$ can

be written out explicitly, as shown in (4.92), with

$$\begin{aligned}
\tilde{\beta}_{\mu}(t,T_0,s) =& \bar{\beta}_{\mu}(t,T_0) + \beta_{\mu}(T_0,s)e^{-\omega_{21}(T_0-t)} \\
\tilde{\beta}_{j,\eta}(t,T_0,s) =& \bar{\beta}_{j,\eta}(t,T_0) + \beta_{\mu}(T_0,s)(e^{-\omega_{2j}(T_0-t)} - e^{-\omega_{21}(T_0-t)}) + \beta_{j,\eta}(T_0,s)e^{-\omega_{2j}(T_0-t)} \\
\tilde{\alpha_{\mu}}(t,T_0,s) =& \beta_{\mu}(T_0,s)[g^{\tilde{Q}}_{2,\mu}(T_0) - e^{-\omega_{21}(T_0-t)}g^{\tilde{Q}}_{2,\mu}(t) - \sum_{j=2}^{3}(e^{-\omega_{2j}(T_0-t)} - e^{-\omega_{21}(T_0-t)})g^{\tilde{Q}}_{j,\eta}(t)] \\
&+\frac{1}{2}\{\beta^2_{\mu}(T_0,s)\mathrm{Var}^{\tilde{Q}}_t[\hat{\mu}(T_0,2)] + \sum_{j=2}^{3}(\beta^2_{j,\eta}(T_0,s)\mathrm{Var}^{\tilde{Q}}_t[\eta_{2,j}(T_0)] \\
&+2\beta_{\mu}(T_0,s)\beta_{j,\eta}(T_0,s)\mathrm{Cov}^{\tilde{Q}}_t[\hat{\mu}(T_0,2),\eta_{2,j}(T_0)])\} \\
&+\sum_{j=2}^{3}\beta_{j,\eta}(T_0,s)[g^{\tilde{Q}}_{j,\eta}(T_0) - e^{-\omega_{2j}(T_0-t)}g^{\tilde{Q}}_{j,\eta}(t)] + \bar{\alpha}_{\mu}(t,T_0) + \alpha_{\mu}(T_0,s). \quad (4.121)
\end{aligned}$$

Bibliography

Adcock, C. (1997). Sample size determination: a review. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 261–283.

Ang, A., D. Papanikolaou, and M. M. Westerfield (2014). Portfolio choice with illiquid assets. *Management Science 60*(11), 2737–2761.

Baran, S., J. Gáll, M. Ispány, and G. Pap (2007). Forecasting hungarian mortality rates using the lee-carter method. *Acta Oeconomica 57*(1), 21–34.

Basak, S. and G. Chabakauri (2010). Dynamic mean-variance asset allocation. *Review of Financial Studies 23*(8), 2970–3016.

Basak, S. and G. Chabakauri (2012). Dynamic hedging in incomplete markets: a simple solution. *Review of financial studies 25*(6), 1845–1896.

Bauer, D., M. Börger, and J. Ruß (2010). On the pricing of longevity-linked securities. *Insurance: Mathematics and Economics 46*(1), 139–149.

Bauer, D., M. Börger, J. Ruß, and H.-J. Zwiesler (2008). The volatility of mortality. *Asia-Pacific Journal of Risk and Insurance 3*(1).

Bauer, D. and J. Ruß (2006). Pricing longevity bonds using implied survival probabilities. In *2006 meeting of the American Risk and Insurance Association (ARIA)*.

Bayraktar, E., M. A. Milevsky, S. David Promislow, and V. R. Young (2009). Valuation of mortality risk via the instantaneous sharpe ratio: applications to life annuities. *Journal of Economic Dynamics and Control 33*(3), 676–691.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science 59*(2), 341–357.

Ben-Tal, A., D. Den Hertog, and J.-P. Vial (2014). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 1–35.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). *Robust optimization*. Princeton University Press.

Blackburn, C. and M. Sherris (2013). Consistent dynamic affine mortality models for longevity risk applications. *Insurance: Mathematics and Economics 53*(1), 64–73.

Blackburn, C. and M. Sherris (2014). Forward mortality modelling of multiple populations. *Working paper*.

Blake, D. and E. Biffs (2012). Keeping some skin in the game: how to start a capital market in longevity risk transfers.

Blake, D., A. J. Cairns, and K. Dowd (2006). Living with mortality: Longevity bonds and other mortality-linked securities. *British Actuarial Journal 12*(01), 153–197.

Booth, H., J. Maindonald, and L. Smith (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population studies 56*(3), 325–336.

Bor, D. and C. Cowling (2013). The ftse 100 and their pension disclosures. Technical report, JLT Pension Capital Strategies.

Börger, M. (2009). The Volatility of Mortality. *Deutscher SCOR-Preis für Aktuarwissenschaften 2008: Zusammenfassung eingereichter Arbeiten 8*, 25.

Börger, M. (2010). Deterministic shock vs. stochastic value-at-risk: an analysis of the Solvency II standard model approach to longevity risk. *Blätter der DGVFM 31*(2), 225–259.

Börger, M., D. Fleischer, and N. Kuksin (2011). Modeling mortality trend under modern solvency regimes. *ASTIN Bulletin*, 1–38.

Börger, M., D. Fleischer, and N. Kuksin (2013, 11). Modeling the mortality trend under modern solvency regimes. *ASTIN Bulletin FirstView*, 1–38.

Brillinger, D. R. (1986). A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. *Biometrics*, 693–734.

Brouhns, N., M. Denuit, and J. Vermunt (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics 31*(3), 373–393.

Cairns, A. J. (2011). Modelling and management of longevity risk: approximations to survivor functions and dynamic hedging. *Insurance: Mathematics and Economics 49*(3), 438–453.

Cairns, A. J. (2013). Robust hedging of longevity risk. *Journal of Risk and Insurance 80*(3), 621–648.

Cairns, A. J., D. Blake, P. Dawson, and K. Dowd (2005). Pricing the risk on longevity bonds. *Life and Pensions 1*(2), 41–44.

Cairns, A. J., D. Blake, and K. Dowd (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance 73*(4), 687–718.

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

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

Cairns, A. J., K. Dowd, D. Blake, and G. D. Coughlan (2014). Longevity hedge effectiveness: A decomposition. *Quantitative Finance 14*(2), 217–235.

Cairns, A. J. B., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin-Actuarial Studies in non LifeInsurance 41*(1), 29.

Caplin, A. and J. Leahy (2006). The recursive approach to time inconsistency. *Journal of Economic Theory 131*(1), 134–156.

Cox, S. H., Y. Lin, R. Tian, and L. F. Zuluaga (2013). Mortality portfolio risk management. *Journal of Risk and Insurance 80*(4), 853–890.

CRO Forum (2010). Longevity risk. *CRO Briefing Emerging Risks Initiative Position Paper, http://www.thecroforum.org/longevity-risk/*.

Czado, C., A. Delwarde, and M. Denuit (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics 36*(3), 260–284.

Dahl, M., S. Glar, and T. Møller (2011). Mixed dynamic and static risk-minimization with an application to survivor swaps. *European Actuarial Journal 1*(2), 233–260.

Dahl, M., M. Melchior, and T. Møller (2008). On systematic mortality risk and risk-minimization with survivor swaps. *Scandinavian Actuarial Journal 2008*(2-3), 114–146.

Damien, P., P. Dellaportas, N. G. Polson, and D. A. Stephens (2013). *Bayesian Theory and Applications*. Oxford University Press.

Dawson, P., K. Dowd, A. J. Cairns, and D. Blake (2010). Survivor derivatives: A consistent pricing framework. *Journal of Risk and Insurance 77*(3), 579–596.

De Jong, F. and P. Santa-Clara (1999). The dynamics of the forward interest rate curve: A formulation with state variables. *Journal of Financial and Quantitative Analysis 34*(01), 131–157.

De Santis, F. (2004). Statistical evidence and sample size determination for bayesian hypothesis testing. *Journal of statistical planning and inference 124*(1), 121–144.

Denuit, M. and A. Goderniaux (2005). Closing and projecting life tables using log-linear models. *Bulletin of the Swiss Association of Actuaries 1*(4), 29–49.

Dowd, K., D. Blake, A. J. Cairns, and P. Dawson (2006). Survivor swaps. *Journal of Risk and Insurance 73*(1), 1–17.

Dowd, K., A. J. Cairns, D. Blake, G. D. Coughlan, and M. Khalaf-Allah (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal 15*(2), 334–356.

Driessen, J., P. Klaassen, and B. Melenberg (2003). The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of Financial and Quantitative Analysis 38*(03), 635–672.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*(6), 1343–1376.

Föllmer, H. and A. Schied (2004). Stochastic Finance, volume 27 of de gruyter studies in mathematics.

Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies 20*(1), 41–81.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

Glasserman, P. and X. Xu (2013). Robust portfolio control with stochastic factor dynamics. *Operations Research 61*(4), 874–893.

Grenadier, S. R. and N. Wang (2007). Investment under uncertainty and time-inconsistent preferences. *Journal of Financial Economics 84*(1), 2–39.

Hansen, L. P. and T. J. Sargent (2001). Robust control and model uncertainty. *American Economic Review*, 60–66.

Hansen, L. P. and T. J. Sargent (2008). *Robustness*. Princeton University Press.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society*, 77–105.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382–401.

Inui, K. and M. Kijima (1998). A markovian framework in multi-factor heath-jarrow-morton models. *Journal of Financial and Quantitative Analysis 33*(03), 423–440.

Jeanblanc, M., M. Yor, and M. Chesney (2009). *Mathematical methods for financial markets*. Springer.

Kallsen, J. (2002). Utility-based derivative pricing in incomplete markets. In *Mathematical FinanceÂ¡Âª Bachelier Congress 2000*, pp. 313–338. Springer.

Karatzas, I. (1991). *Brownian motion and stochastic calculus*, Volume 113. springer.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the american statistical association 90*(430), 773–795.

Kogure, A., K. Kitsukawa, and Y. Kurachi (2009). A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in Japan. *Asia-Pacific Journal of Risk and Insurance 3*(2).

Kogure, A. and Y. Kurachi (2010). A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics 46*(1), 162–172.

Lee, R. and T. Miller (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography 38*(4), 537–549.

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting us mortality. *Journal of the American statistical association 87*(419), 659–671.

Li, H. (2014). Dynamic hedging of longevity risk: the effect of trading frequency. *Available at SSRN 2532422*.

Li, H., A. De Waegenaere, and B. Melenberg (2013). The choice of sample size for mortality forecasting: A Bayesian learning approach.

Li, H., A. De Wagenaere, and B. Melenberg (2014). Robust longevity risk management.

Li, J. S.-H. and M. R. Hardy (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal 15*(2), 177–200.

Li, J. S.-H. and A. Luo (2012). Key q-duration: A framework for hedging longevity risk. *Astin Bulletin 42*(2), 413–452.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography 42*(3), 575–594.

Lin, Y. and S. H. Cox (2005). Securitization of mortality risks in life annuities. *Journal of risk and Insurance 72*(2), 227–252.

Lin, Y. and S. H. Cox (2008). Securitization of catastrophe mortality risks. *Insurance: Mathematics and Economics 42*(2), 628–637.

Mania, M., M. Schweizer, et al. (2005). Dynamic exponential utility indifference valuation. *The Annals of Applied Probability 15*(3), 2113–2143.

Menoncin, F. (2008). The role of longevity bonds in optimal portfolios. *Insurance: Mathematics and Economics 42*(1), 343–358.

Miltersen, K. R. and S.-A. Persson (2005). Is mortality dead? stochastic forward force of mortality rate determined by no arbitrage. Technical report, Working Paper, University of Bergen.

Mulvey, J. M., R. J. Vanderbei, and S. A. Zenios (1995). Robust optimization of large-scale systems. *Operations research 43*(2), 264–281.

OECD (2013). Pension at a glance 2013. *Techical report, http://dx.doi.org/10.1787/pension_glance-2013-en*.

O'Hare, C. and Y. Li (2012). Identifying structural breaks in stochastic mortality models. *Available at SSRN 2192208*.

Pedroza, C. (2002). *Bayesian Hierarchical Time Series Modeling of Mortality Rates*. Ph. D. thesis, Havard University.

Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics 7*(4), 530–550.

Pitacco, E., M. Denuit, S. Haberman, and A. Olivieri (2009). *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press, USA.

Plat, R. (2009). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics 45*(1), 123–132.

Plat, R. (2011). One-year value-at-risk for longevity and mortality. *Insurance: Mathematics and Economics 49*(3), 462–470.

Reichmuth, W. H. and S. Sarferaz (2008). Bayesian demographic modeling and forecasting: an application to US mortality. Technical report, SFB 649 discussion paper.

Reid, M. D. and R. C. Williamson (2011). Information, divergence and risk for binary experiments. *The Journal of Machine Learning Research 12*, 731–817.

Renshaw, A. E. and S. Haberman (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics 38*(3), 556–570.

Rockafellar, R. T. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance 26*(7), 1443–1471.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 165–180.

van Berkum, F., K. Antonio, M. Vellekoop, and K. Leuven (2013). Structural changes in mortality rates.

Wang, F. and A. E. Gelfand (2002). A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. *Statistical Science 17*(2), 193–208.

Wang, S. (2002). A universal framework for pricing financial and insurance risks. *Astin Bulletin 32*(2), 213–234.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *Journal of the Royal Statistical Society: Series D (The Statistician) 46*(2), 185–191.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. Springer.

Wong, T. W., M. C. Chiu, and H. Y. Wong (2014). Time-consistent mean–variance hedging of longevity risk: Effect of cointegration. *Insurance: Mathematics and Economics 56*, 56–67.

Zhu, S. and M. Fukushima (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research 57*(5), 1155–1168.