Modeling mortality

Hari, N.

Publication date:
2007

Document Version
Peer reviewed version

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Hari, N. (2007). *Modeling mortality: Empirical studies on the effect of mortality on annuity markets*. CentER, Center for Economic Research.

Modeling Mortality:
Empirical Studies on the Effect of Mortality on Annuity Markets

Modeling Mortality: Empirical Studies on the Effect of Mortality on Annuity Markets

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof. dr. F.A. van der Duyn Schouten, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 12 januari 2007 om 10.15 uur door

Norbert Hári

geboren op 2 april 1977 te Celldömölk, Hongarije.

Promotores:	prof. dr. Bertrand Melenberg
	prof. dr. Theo E. Nijman
Copromotor:	dr. Anja De Waegenaere

Acknowledgements

When I came to Tilburg for the first time in August 1999, I had never imagined that one day I would defend my PhD thesis here.

I would like to express my gratitude towards those who directly and indirectly helped me in the past four years in the process of writing of this thesis. My special thanks go to Anja De Waegeneare, Bertrand Melenberg and Theo Nijman, who have been my supervisors during the entire PhD project. They motivated me continuously, and provided stimulating advice and expertise during our discussions. I am grateful for their comments, ideas and tireless advocacy on rephrasing during writing the thesis.

I am very grateful to my dissertation committee: Bas Werker, Frank de Jong, Natacha Brouhns and Johan Mackenbach, for taking the time to read my thesis. It is an honor to have them in my committee.

I also thank the Finance Department for creating a stimulating research environment and for organizing several amazing social events. I am sincerely grateful to Steven Ongena for his support. I appreciate the patience and company of my office mate Ralph. My life in Tilburg would not have been so enjoyable without my friends and colleagues: Attila, Ákos, Péter, Igor, Marta, Chendi, Marina, Zoli and Jutka, and many others.

My deepest gratitude goes to my wife. It would not have been possible to complete my doctoral studies without the continuous and unconditional support of Kriszti. She provided me with love and inspiration, and tolerated my workaholic life style and helped me to cope with the challenge day by day. I dedicate this thesis to her.

Last but not least, I would like to thank my parents, my brother Balázs, my parents-in-law and other relatives for their support and the many joyful moments we shared.

Thank you all.

Tilburg, August 2006

Table of Contents

1 Introduction **1**

1.1 Patterns in survival rates 2

1.2 Motivation and overview of the thesis 9

2 Literature Survey **13**

2.1 Cross-sectional models on human mortality 13

2.2 Dynamic models on human mortality 15

2.3 Longevity risk in mortality projections 22

2.4 Heterogeneity in survival rates 24

2.5 Contribution of the thesis 25

3 Estimating the Term Structure of Mortality **27**

3.1 Introduction 27

3.2 The Lee and Carter approach 29

3.3 Lee-Carter with time-varying drift 34

3.4 Empirical analysis 37

3.4.1 Data 37

3.4.2 Further specifications 38

3.4.3 Sample period sensitivity 38

3.4.4 Estimation results 41

3.4.5 Prediction 47

3.5 Summary and conclusion 50

3.A Appendix on spline interpolation 52

4 Longevity Risk in Portfolios of Pension Annuities **55**

4.1 Introduction 55

4.2 Forecasting future mortality 58

4.3 Uncertainty in expected remaining lifetime 61

4.4 Effect of longevity on market value of annuities 63
4.5 Effect of longevity risk on funding ratio uncertainty 65
4.5.1 Fund characteristics 65
4.5.2 Market value of assets and liabilities 66
4.5.3 The funding ratio distribution 69
4.6 Management of longevity risk 73
4.6.1 Calibrating the solvency buffer 74
4.6.2 Pricing reinsurance contracts 76
4.6.3 Effect on funding ratio distribution 77
4.7 Effect of combined longevity and market risk 77
4.7.1 Data 79
4.7.2 Uncertainty in the future funding ratio 79
4.8 Conclusions 82
4.A Parameter estimates of the mortality model 84
4.B Age and gender distribution of the Dutch population 86
4.C Projection method 87

5 The Determinants of the Money's Worth of Participation in Collective Pension Schemes 91
5.1 Introduction 91
5.2 A survey of the literature on money's worth of annuities 96
5.3 The money's worth of participation in collective pension schemes 98
5.4 The money's worth in conditionally indexed collective pension schemes . 106
5.5 Conclusions 110
5.A Socioeconomic life tables 112
5.B Modeling survival probabilities 113
5.C Money's worth of participation in collective pension schemes 117

6 Conclusions and Directions for Further Research 123
6.1 Summary and conclusions 123
6.2 Some directions for further research 124

References 127

Nederlandse Samenvatting (Dutch Summary) 137

Chapter 1

Introduction

The population of many countries might undergo dramatic changes in the coming decades due to continuous increases in life expectancy. The fact that people seem to live longer and the low fertility rates contribute to an increasing share of elderly people in the total population in the future. Carone et al. (2005) discuss the macroeconomic aspects of ageing, such as the impact on productivity, labor supply, capital intensity, employment and economic growth, and the indirect effects on the economy via budgetary effects. We analyze a subset of these issues, with the aim to shed light on the interaction between ageing and the financial markets. The thesis considers the implication of longevity and related risks on the value of financial instruments linked to human survival, such as life annuities.

This chapter gives some basic statistics and describes the stylized facts found in historical survival data. An overview of the historical evolution of life expectancy is presented for selected European countries, with an emphasis on the Netherlands. Although from the perspective of a pension provider, life expectancy at adult and elderly ages might be more relevant than life expectancy at birth, this chapter also devotes attention to the young, because one of the later chapters models the survival characteristics of the full age spectrum. Projected life expectancies in Europe based on the study of the Economic Policy Committee and European Commission (2005) are considered. Heterogeneity in expected lifetime among various socioeconomic groups is also discussed.

1.1 Patterns in survival rates

Carone et al. (2005) report that life expectancy at birth increased by 8 years between 1960 and 2000 in Europe. Based on projections[1] of the Economic Policy Committee and European Commission (2005), life expectancy at birth is projected to increase by 6 years for men and by more than 5 years for women till 2050 in the populations of the 25 member countries of the EU. This means, that if the projections are right and the fertility remains at the current level, the share of the elderly in the total population is going to increase in the future, creating a potentially higher pressure on the social security systems. Carone et al. (2005) claim that much of the gain in life expectancy is expected to stem from lower mortality rates of the elderly, because the life expectancy at the age of 65 is predicted to increase by almost 4 years between 2004 and 2050. Table 1.1 gives life expectancy at birth in historical years for selected countries based on the data of the Human Mortality Database (HMD)[2], and projected life expectancy at birth in 2050 based on the report of the Economic Policy Committee and European Commission (2005).

In the second half of the 20th century there is a clear pattern of increasing life expectancy for all countries. First, the expected lifetime at birth shows a more than 10-year improvement in the case of Austria, Belgium, Finland, and France between 1950 and 2000. Spain experienced an improvement of more than 15 years. Most of these increases were likely due to medical advances and better standard of living. The speed of improvement in Hungary seemed to slow down in the last quarter of the previous century. In terms of the projections, there seems to be convergence[3] of expected remaining lifetimes among countries and genders in the first half of the 21st century. The fact that expected lifetime of women is higher than for men in all countries is also present in the table, and this finding remains valid in other historical years not shown in the table.

[1] The main assumptions behind the forecasts of the age-specific mortality rates are as follows: 1. The trends of decreasing age-specific mortality rates observed over the period 1985 to 2002 continue between 2002 and 2018. 2. The decreasing trends slow down between 2018 and 2050. 3. The forecasts incorporated additional assumptions on the convergence of life expectancy at birth among the EU10 and EU15 Member States.

[2] Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 04.11.2005).

[3] This result highly depends on the assumptions of the model used to produce projections. The report by the Economic Policy Committee and European Commission (2005) assumes the convergence of life expectancy at birth among EU countries.

	Men				Women			
	1950	1975	2000	2050	1950	1975	2000	2050
Austria	62.2	67.7	75.0	82.8	67.3	74.7	81.1	87.2
Belgium	63.8	68.8	74.6	82.1	68.9	75.2	80.9	87.5
Czech Republic	62.0	67.0	71.6	79.7	66.8	74.0	78.3	84.1
Denmark	69.1	71.3	74.4	81.4	71.5	77.0	79.1	85.2
UK	66.5*	69.7*	75.7*	82.4	71.3*	75.9*	80.4*	86.7
Finland	60.4	67.4	74.2	81.9	67.9	76.1	81.0	86.6
France	63.4	69.0	75.3	82.3	69.2	76.9	82.8	87.9
Germany	n/a	68.1**	75.3**	82.0	n/a	74.7**	81.2**	86.8
Hungary	59.9	66.3	67.4	78.1	64.3	72.4	76.0	83.4
Italy	64.0	69.5	76.6	82.8	67.5	75.9	82.5	87.8
Netherlands	70.3	71.5	75.7	81.1	72.6	77.7	80.8	85.2
Spain	59.4	70.5	75.8	81.7	64.2	76.3	82.7	87.3
Sweden	69.8	72.2	77.4	82.6	72.4	77.9	82.0	86.6

* England and Wales
** West Germany

Table 1.1: Historical and forecasted life expectancy at birth, in years. The table presents gender-specific historical and forecasted expected lifetime at birth in selected EU countries. The historical life expectancies are provided by the Human Mortality Database, the forecasted ones are based on the report by the Economic Policy Committee and European Commission (2005).

For risk management or pricing purposes it is crucial to know whether either the improvement affects all the age groups equally, or whether the survival chances of some groups increased more than for others. The age group of 65 receives large attention from pension providers, since the time spent in retirement crucially influences the calculation of contributions. Table 1.2 gives the expected remaining lifetime of men and women with age 65 in the same set of countries as in Table 1.1.

Table 1.2 shows the substantial improvement in life expectancy also for the 65-year-old. Most of the gain - more than 1 year per decade - was realized in the last quarter of the 20th century both for men and women, in almost all of the selected countries. The projections of the Economic Policy Committee and European Commission (2005) seem to suggest that the improvement in the expected lifetime of the elderly will continue in the future.

Tables 1.1 and 1.2 suggest that the improvement in survival chances affected both the young and the elderly. In order to explore the size of the improvement for the age groups in more detail, I illustrate some stylized facts based on the Human Mortality Database. Since all the work in the thesis is based on the mortality data of the Netherlands, I first illustrate the historical mortality experience of the Dutch total population. Figure 1.1 displays the survival function $l_{x,t}$, which gives the number of expected survivors from birth to exact age x in calendar year t (e.g. in 2000, etc.), where the number of expected

	Men				Women			
	1950	1975	2000	2050	1950	1975	2000	2050
Austria	12.1	12.2	15.9	20.4	13.7	15.6	19.5	23.6
Belgium	12.4	12.2	15.5	20.3	14.0	15.7	19.6	24.1
Czech Republic	11.7	11.3	13.6	18.4	13.3	14.6	17.1	20.9
Denmark	13.6	13.8	15.2	19.3	14.1	17.3	18.2	21.9
UK	11.9*	12.4*	15.8*	20.4	14.4*	16.5*	19.1*	23.3
Finland	10.9	12.1	15.5	20.0	13.1	15.9	19.4	23.3
France	12.2	13.2	16.7	20.5	14.6	17.2	21.2	24.5
Germany	n/a	12.1**	15.8**	20.1	n/a	15.5**	19.5**	23.4
Hungary	12.5	12.0	12.8	18.6	13.6	14.6	16.6	21.1
Italy	13.3	13.1	16.6	20.4	14.3	16.3	20.5	24.1
Netherlands	14.1	13.5	15.4	18.9	14.6	17.2	19.4	22.1
Spain	12.3	13.6	16.6	20.0	14.3	16.6	20.6	23.7
Sweden	13.5	14.1	16.7	20.0	14.3	17.3	20.1	23.0

* England and Wales
** West Germany

Table 1.2: Historical and forecasted life expectancy conditional on having reached the age 65, in years. The table presents gender-specific historical and forecasted expected lifetime at the age of 65 in selected EU countries. The historical life expectancies are provided by the Human Mortality Database, the forecasted ones are based on the Economic Policy Committee and European Commission (2005) report.

survivors is calculated based on the assumption that age-specific survival characteristics prevailing in period t also hold for any other (historical and future) time period[4]. The initial size of the cohorts at birth is normalized to 100,000 people.

Similarly, death curves $d_{x,t}$ which plot the expected number of people in a cohort with age between x and $x+1$ dying during year t are also based on the survival characteristics corresponding with period t, assuming that survival chances do not change over time (Figure 1.2). The total size of the cohort at birth is again normalized to 100,000. Survival functions and death curves at time t are interrelated, because they represent the age-specific survival characteristics of the same reference population[5].

Both figures (Figures 1.1 and 1.2) show that survival characteristics were changing over time, because we got various curves in different historical years. The changing shape of the curves representing the mortality as a function of the attained age shows two stylized facts in the data (Olivieri, 2001). First, there is an increasing concentration of deaths around the mode (at old ages) of the curve of deaths over time, which is also reflected in the "rectangularization" of the survival function. This reflects that fewer

[4]For instance, the number of survivors in a cohort aged 25 in year 2000 (or equivalently, the size of the cohort aged 25 born in 1975) is calculated based on the assumption that the survival chances between 1975 and 2000 are the ones observed in 2000.

[5]More precisely, $d_{x,t} = l_{x,t} - l_{x+1,t}$

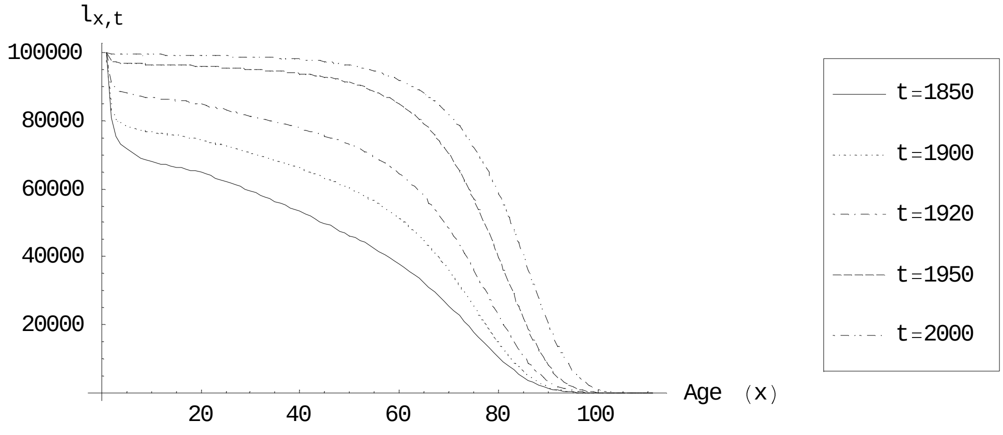

Figure 1.1: Survival functions. The figure plots the survival functions defined as the number of expected survivors in a cohort aged x at calendar year t with an initial cohort size of 100,000 at birth, where the survival characteristics correspond with period t. Data source: Human Mortality Database.

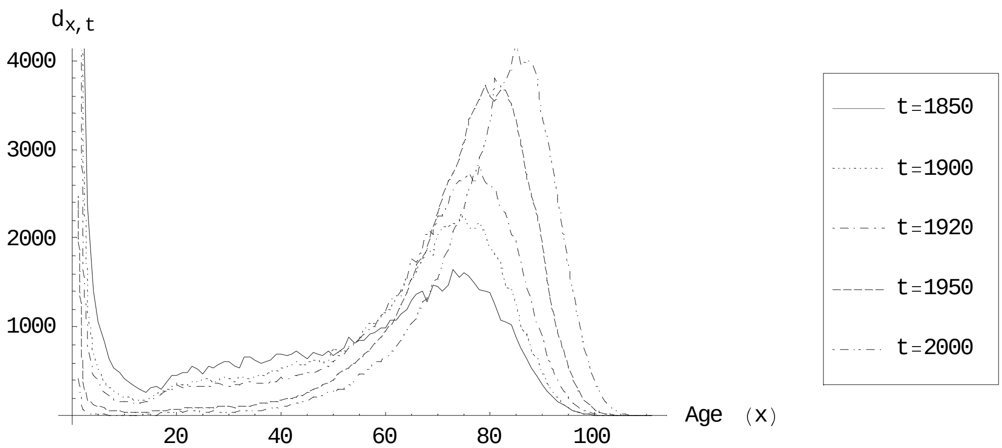

Figure 1.2: Death curves. The figure plots the expected number of people dying in a cohort aged between x and $x+1$ for selected calendar year t with an initial cohort size of 100,000 at birth, where the survival characteristics correspond with period t. Data source: Human Mortality Database.

people die at the young and adult age, and elderly people tend to die in an age interval which is getting narrower. The second phenomenon is the "expansion" of the survival function, characterized by the shift of the death curve and the survival function towards very old ages, implying that the maximum attainable age also shifted upwards.

The above figures already give insight into the change of survival prospects over time for all the age groups. The graph constructed for plotting the time evolution of the mortality rate of an age group conditional on attaining a specific age is called the "mortality profile", and gives a more precise representation of differences between the historical mortality evolutions of age groups. Figures 1.3 and 1.4 show the evolution of the mortality profile $q_{x,t}$ (the probability of dying[6] during year t conditional on having reached age x) of groups with different ages in the total Dutch population between 1850 and 2003.

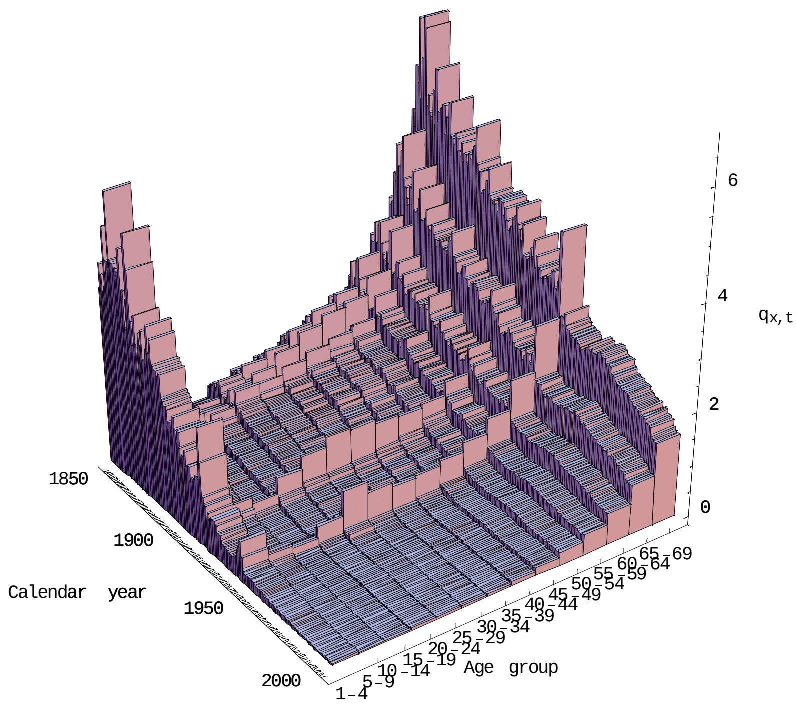

Figure 1.3: Mortality profile for the young and adult (1-69 years). Historical evolution of age-specific probability of death for the total Dutch population. Source of data: Human Mortality Database.

[6]We refer the reader to Gerber (1997) for the details on estimating $q_{x,t}$.

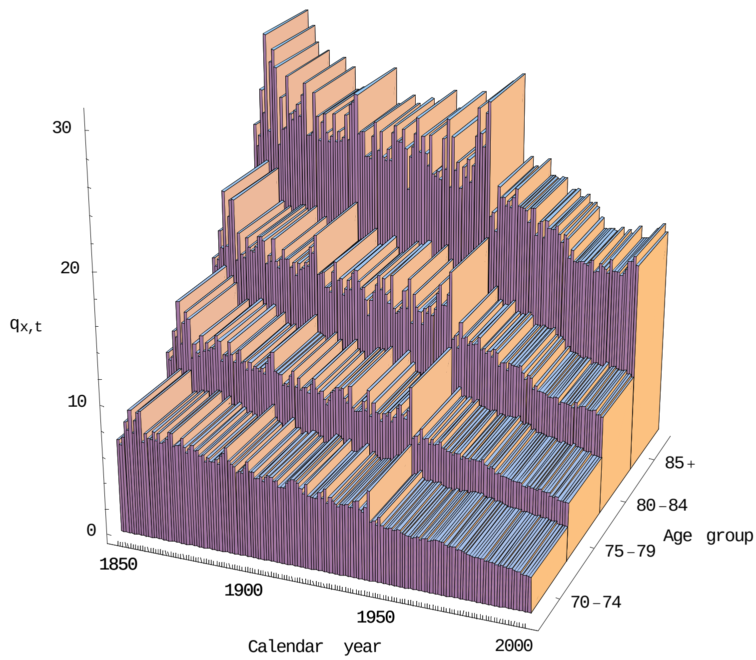

Figure 1.4: Mortality profile for the elderly (70+ years). Historical evolution of age-specific probability of death for the total Dutch population. Source of data: Human Mortality Database.

From Figures 1.3 and 1.4 it is clear that there was a highly volatile period at the beginning of the sample and two peaks in the first half of the 20th century. The first peak in 1918 is related to the outbreak of the so-called "Spanish flu" epidemic, while the second one is due to the "Dutch Hunger Winter" in 1944-45. The mortality profiles show a decline in the 1-year conditional death probabilities for all age groups. Figure 1.3 shows the remarkable decline in mortality of the youngest age groups between 1850 and 1950, while the improvement in mortality of the young and middle aged was less spectacular, but still important. After the 1950-s, the mortality of the young and adult population reached a very low and stable level. Figure 1.4 illustrates that mortality rates of the elderly were decreasing in the last 150 years, and the rate of decrease did not slow down at the end of the sample period. Moreover, as Carone et al. (2005) conjecture, it is very likely that a large part of the gain in expected lifetime is going to be attributed to the increasing survival probabilities[7] of the elderly in the future.

[7]When I mention survival probabilities, I explicitly refer to the probability of surviving during year

Another important aspect of the historical plot of the mortality profiles is the time variation of mortality rates in the past. Before 1950, the variation in mortality rates in the young and adult age groups is much larger than after 1950. There seems to be some decrease in variability after 1950 for the elderly as well, but the pattern of mortality does not decrease as smoothly as for younger groups. The above figures clearly show the time variation in historical human mortality rates around the decreasing trend. If we assume that the variability in mortality rates experienced in the past is also going to be reflected in the future behavior of death probabilities, then the question arises, whether this is an important risk component of the overall riskiness in the portfolio of companies selling survival related instruments.

Apart from the heterogeneity in survival chances among age groups, we already saw in Tables 1.1 and 1.2 that women have longer expected lifetime than men. However, there are lots of other characteristics which make individuals differ from each other even at the same age and in the same gender group.

There are well observed factors documented in the finance[8] literature which signal heterogeneity in survival. For instance, Kunst (1997) found the effect of different educational levels on life expectancy in several European countries. Huisman et al. (2004, 2005) also documented mortality differences among cohorts with different educational levels in European populations. Mackenbach et al. (2003) find that the differences in socioeconomic inequality related mortality were widening between 1983 and 1993. A report by Herten et al. (2002) documents heterogeneity in survival rates along educational lines in the Netherlands, illustrated in Table 1.3.

Based on a social economic survey between 1995 and 1999, Table 1.3 shows that women with average education at the age of 20 are expected to live 5.4 years longer than men. This difference between women and men slightly decreases to 4.7 for people who attained the age of 65. The difference in expected lifetime is present among cohorts with different educational level. 20-year-old high educated men are expected to live 5 years longer than the ones with the lowest education. This difference shrinks to 3.7 years as soon as men reach the age of 65. A 20-year-old woman with high education lives 2.6 years longer in expectation than a woman with the lowest education, and this difference becomes 2.1 years as a woman gets 65 years old.

---

t, conditional on having reached age x, denoted by $p_{x,t}$. The relationship between (1-year) survival and death probabilities is the following: $p_{x,t} = 1 - q_{x,t}$.

[8]We do not take into account explicitly health related issues like diseases, drinking, and smoking habits etc., which directly influence the survival chances of people. These factors are discussed in more detail in the medical literature.

	Low Education	Lower Secondary Education	Higher Secondary Education	High Education	Difference High-Low Education	Average
	Men					
0 yrs	73.1	76.0	76.0	78.0	4.9	75.0
20 yrs	53.6	56.5	56.5	58.5	5.0	55.5
65 yrs	11.1	13.4	13.3	14.8	3.7	12.4
	Women					
0 yrs	79.5	82.0	82.1	82.1	2.6	80.5
20 yrs	59.9	62.4	62.5	62.5	2.6	60.9
65 yrs	16.4	18.5	18.6	18.5	2.1	17.1

Source: Herten et al. (2002), Table 1.

Table 1.3: Heterogeneity in expected lifetime. The table gives the expected remaining lifetime for the newly born and for people conditional on having reached the age of 20 and 65, grouped along different educational backgrounds and gender, in the Netherlands.

Besides the differences in educational level, gender or age, other characteristics, such as different area of living (rural / urban areas), or ethnicity etc., also make mortality rates vary (see for instance Bos et al., 2005).

1.2 Motivation and overview of the thesis

If future probabilities of survival were known with certainty, the expected lifetime and, therefore, the expected number of people dying in a given year would also be known with certainty. However, the lifetime of an individual and the realized number of deaths in a pool are uncertain ex ante. This risk is called *micro-longevity risk* throughout the thesis. In an infinitely large pool, on average, people "die according to expectation" (Law of Large Numbers)[9]. This implies that increasing the number of participants in a pool will decrease the relative size of micro-longevity risk to zero. However, as the data presented in the previous section already suggested, survival probabilities in the future are far from certain. This creates an additional source of uncertainty, called *macro-longevity risk*, which cannot be reduced by increasing the number of the policyholders in a pool. In order to measure and deal with this risk, we need to model and forecast survival probabilities, which is going to be the central theme of this thesis. When

[9]The Law of Large Numbers states that in an infinitely large pool where the lifetime of the members is independently and identically distributed, the sample average converges almost surely to the common expected value.

the contribution rates of the policyholders of a given life annuity contract or the price of a life insurance contract are calculated, the uncertainty around the forecasts has to be incorporated into the prices. For instance, pension funds or annuity providers are exposed to a substantial amount of loss if the survival prospects of the existing policyholders improve significantly and the effect of the realized improvement was not incorporated in the pricing and reserving calculations. On the contrary, life insurers face the risk of unexpected drop in future survival rates.

The possible consequences of macro-longevity risk received large attention particularly in the year 2000, when the Equitable Life Assurance Society (ELAS) failed due to the exposure to both interest rate risk and (to a lesser extent) macro-longevity risk, and was closed to new business (Blake et al., 2006). ELAS sold (with profit) pension annuities with guaranteed annuity rates, and the pricing was based on specific assumptions regarding to future interest and mortality rates. However, the lower than expected interest rates and higher than expected life expectancy made the annuities very valuable. Besides the poor state of interest rate management of ELAS, the significant exposure to longevity risk led to the acknowledgement that mortality risk is a key risk factor, which cannot be ignored.

If a pension annuitant is shorter lived in expectation than the average individual in the annuity group, and the annuities are priced to reflect the longevity of the average individual in the group, then the price of the annuity ex ante will not be actuarially fair from the standpoint of the shorter lived individual. If the contract was fairly priced, how much more should a longer than average lifetime individual pay for a life annuity contract than an individual with a shorter than average lifetime, if they are at the same age? Does heterogeneity in expected lifetime translate to sizeable differences in fair value? Do the differences stimulate incentives not to buy the insurance/annuity contract, or do they create adverse selection? If these differences are substantial, the shorter than average lived individuals have the incentive to opt out of the pension annuity market. Typically, individuals with higher than average expected lifetime self-select into a life annuity contract, which are priced based on the average characteristics of the population. This type of adverse selection leads to overpricing of annuity contracts. Finkelstein and Poterba (2004) report that on the voluntary markets of the US and the UK, the expected present discounted value of payout to a typical individual is only 80 to 85 percent of the annuity premium. Part of the difference is due to the administrative costs, but roughly half appears to be due to adverse selection. The asymmetric information between insureds and pension funds make the contracts more expensive, which implies that the effect of heterogeneity is potentially important. Even though compulsory/collective systems mitigate adverse selection, the differences in the actuarially fair price of life insurance or life annuity contracts due to heterogeneity are not resolved.

The following chapters of the thesis are organized as follows. Chapter 2 gives an overview of the literature related to mortality projection and the inherent longevity risk. Nowadays, several classes of models exist. I only discuss classes which I think help the reader to understand the basic facts in mortality modeling and that are inevitable to understand the main ideas behind the model developed in a later chapter. Starting from the seminal contributions in the 19th century, I give a short description of the models which contributed to modern mortality modeling to a large extent, with a distinct emphasis on the most recent literature. Then, I will address some of the papers of the so-called money's worth literature, which discusses the effect of survival heterogeneity on the expected present value of annuity payment per the amount spent to purchase the annuity.

Chapter 3 introduces a model for human mortality rates. In the benchmark methodology (Lee and Carter, 1992), the time variation of the age-specific log mortality rates is explained by a linear combination of factor(s). In this chapter we formulate a generalized model starting from the benchmark. We estimate various specifications of the generalized model, and illustrate them by forecasting age-specific mortality rates with the related prediction intervals by using Dutch mortality data.

Chapter 4 analyzes the importance of micro- and macro-longevity risk for the solvency position of representative pension funds of various sizes. We use the estimates of Chapter 3 and assess the importance of uncertain future survival probabilities. First, we analyze the effect of longevity risk on the funding ratio[10] of pension funds by assuming no financial risk (for example, interest rate, stock market return risk). We calibrate the minimum size of the initial funding ratio by taking into account various sources of longevity risk in order to decrease the probability of insolvency to a very low level for several time horizons. Second, we investigate the relative importance of longevity risk with the presence of market risk.

Chapter 5 measures the present value of a single year participation in a pension scheme consisting of heterogeneous participants, where the participation in the scheme is compulsory. In many countries, the contributions to such schemes are often set uniformly (the same percentage of the salary for all the participants), irrespective of age, gender,

[10]The funding ratio at time t is the ratio of the market value of assets and liabilities. We call a fund solvent at time t, if the funding ratio of the fund is greater than or equal to 1.

or education level. We quantify the effect of survival heterogeneity on the fair price of participation and the incentives which arise due to uniform pricing are going to be addressed. We investigate nominal, real, and indexed pension schemes.[11]

Chapter 6 concludes and provides possible directions for further research.

[11] In a nominal scheme, the future benefits are defined in nominal terms, while in a real scheme they are defined in real terms (future inflation does not deteriorate the purchasing power). Indexed schemes provide no, partial, or full indexation against future inflation, depending of the future funding ratio of the fund.

Chapter 2

Literature Survey

Epidemiological factors seem to have contributed substantially to the increase in life expectancy through prevention of diseases as an important cause of mortality at younger ages. Vaccination and antibiotics together with the improved living standards seem to have increased the life expectancy even further, and chronic diseases became the leading cause of death in most of the developed countries. However, we do not discuss cause-specific mortality models in detail due to the following reason. Pension funds or insurance companies are much more interested in "all-cause" mortality, because the total cost of a plan does not change with changes in the causes of death unless the compositions of cause-specific mortality add up to different totals (Girosi and King, 2005a). Furthermore, they are much more interested in few common risk factors, which replace all the known and unknown factors (if they exist) that drive the total mortality of the policyholders. The few factors are intended to reproduce the variability of mortality rates with a potentially small information loss. Forecasts based on a limited number of factors are more reliable due to the fewer parameters which need to be estimated, compared to the cause-specific mortality models which cover the full spectrum of epidemiological mortality risks. The next sections will give an overview on purely statistical models in both the descriptive and predictive sense. Almost all exclude exogenous demographic and epidemiological risk factors.

2.1 Cross-sectional models on human mortality

The literature first concentrated on describing the cross-section of death probabilities, with the primary objective to smooth data, to eliminate/reduce errors, to create life tables or to add inferences to incomplete data (Keyfitz, 1982). Parameterization functions are often called mortality 'laws' and they describe mortality age patterns in terms of functions of age. The so called 'Gompertz law' (Gompertz, 1825), or 'Makeham's law' (Makeham, 1860) are among the earliest examples of formulae adopted for mortality modeling purposes. According to the Gompertz law the force of mortality[1] (μ_x) of a person aged x is modeled as follows:

$$\mu_x = B \exp\left[\theta x\right], \tag{2.1}$$

and according to the Makeham's law:

$$\mu_x = A + B \exp\left[\theta x\right], \tag{2.2}$$

with A, B, and θ unknown parameters. The constant A which is an additional component in (2.2) can be thought of as representing the risk of death which is independent of age, and the exponential term is responsible for capturing the differences in mortality across ages.

The Gompertz-Makeham curves were further developed. For instance, Perks (1932) modified (2.2):

$$\mu_x = \frac{A + B \exp\left[\theta x\right]}{1 + C \exp\left[\theta x\right]}. \tag{2.3}$$

This functional form allows one to fit the slower rate of increase in mortality at older ages, since mortality levels off at advanced ages.

The second group of models are the additive multi-component models. Due to the differences in the factors driving the mortality of different parts of the mortality curve, a model for the force of mortality as a function of three components was developed by Thiele (1872). Thiele claimed that the cause of death falls into one of three classes: one component represents the mortality at infancy and childhood, the second one is responsible for capturing the mortality behavior for the adulthood, and the last component describes the mortality of the elderly. The sum of the components describes the mortality pattern across the entire age span:

$$\mu_x = A_1 \exp\left[-B_1 x\right] + A_2 \exp\left[-\frac{1}{2} B_2 (x - c)^2\right] + A_3 \exp\left[-B_3 x\right]. \tag{2.4}$$

[1] Or in other words, the force of mortality is called the instantaneous probability of death: the hazard rate that a person aged x does not survive age $x + \triangle t$, where $\triangle t$ is infinitesimally small. The force of mortality can be estimated as follows: $\widehat{\mu}_x = D_x / E_x$. D_x is the number of people with age x that died in a given year, and E_x is the exposure being the number of person years with age x in the same year. For more details, see e.g. Gerber (1997).

Heligman and Pollard (1980) proposed a model with three components which are analogous to the ones in Thiele's model. It also captures the mortality curve over the entire age range and it is called the Heligman-Pollard law:

$$q_x = A^{(x+B)^C} + D \exp\left[-E(\ln x - \ln F)^2\right] + \frac{GH^x}{1+GH^x}, \tag{2.5}$$

where q_x denotes the conditional 1-year probability of death of an individual aged x.

Polynomial models became popular, because most mortality curves can be approximated by a polynomial with high accuracy. However, if (high-order) polynomials are extended far beyond the age range from which they are estimated, they are susceptible to produce unpredictable shapes. In most of the cases, the extended mortality curves do not match the expected behavior of mortality curves (mortality rates become negative for the elderly etc.), because the shape of polynomials can be arbitrary outside the data range. Mortality laws do not suffer from this weakness; however, they do not have such a perfect fit in the sample.

Alternatively, mortality laws were combined with polynomial techniques. For instance, in order to increase the fit of the mortality curves, the Gompertz-Makeham mortality law was combined with polynomials to any degree (Forfar et al., 1988; Sithole et al., 2000):

$$\mu_x = \sum_{i=1}^{r-1} \alpha_i x^i + \exp\left[\sum_{j=0}^{s-1} \beta_j x^j\right]. \tag{2.6}$$

2.2 Dynamic models on human mortality

The mortality curves which plot the 1-year conditional death probabilities as a function of the attained age are time-varying, as it was illustrated in Chapter 1. Several studies fitted a cross-sectional model on the time series of mortality data, and the time series of the fitted parameters were used to forecast future mortality by ARIMA modeling (McNown and Rogers, 1989). Tabeau (2001) claims that annual estimates of model parameters are rather unstable, so that a joint model for subsequent mortality data in the form of two dimensional mortality surfaces is necessary: mortality has to be modeled as a function of age and time. Already the early seminal contributions estimated models with a dynamic nature, where the age pattern of mortality deterministically depends on the calendar year via the parameters of the model. For instance, Blaschke (1923) estimated a dynamic Makeham's law. In an alternative approach, various parameters of the Heligman-Pollard law are extended to be a function of the calendar year, see, for instance, Heligman and Pollard (1980) and Benjamin and Soliman (1993). Making future predictions is done by assigning values to the two predictors, time and age.

Renshaw et al. (1996) generalized the polynomial models via higher order polynomials as a function of age and time. They estimated polynomials which fit mortality in both time (t) and age (x) simultaneously:

$$\mu_{xt} = \beta_0 + \sum_{j=1}^{s} b_i L_j(x) + \sum_{i=1}^{r} \alpha_i t^j + \sum_{i=1}^{r}\sum_{j=0}^{s} \gamma_{ij} L_j(x) t^j, \tag{2.7}$$

where $L_j(x)$ represents an orthonormal (Legendre-)polynomial of degree j, and β_0, b_i, α_i and γ_{ij} are unknown parameters. By increasing the order of the polynomials, the model can fit the data extremely well. However high-order polynomials are susceptible to produce unpredictable shapes when used to extrapolate beyond the original data[2].

In the 1990-s, a new model for forecasting the age pattern was proposed by Lee and Carter (1992), which allows for uncertainty in projected rates via a stochastic process driving the log mortality rates and capturing the period effects. Age-specific log mortality rates are constructed by an affine transformation in terms of the sum of a time-invariant age-specific constant (α_x) and a product of a time-varying single latent factor (γ_t) and an age-specific time-invariant component (β_x). The resulting model equals

$$m_{x,t} = \alpha_x + \beta_x \gamma_t + \delta_{x,t}, \tag{2.8}$$

where $m_{x,t}$ denotes the log central death rate of a person with age $x \in \{1, ..., na\}$, and at time $t \in \{1, ..., T\}$.[3] α_x describes the average age-specific pattern of mortality, γ_t represents the general mortality level, and β_x captures the age-specific sensitivity of individual age groups to the general level of mortality changes. $\delta_{x,t}$ is the age- and time-specific innovation term, which is assumed to be a white noise, with zero mean.

The model in (2.8) is not identified, since the distribution is invariant with respect to the following parameter transformations. If $\alpha = (\alpha_{x_1}, ..., \alpha_{x_{na}})'$, $\beta = (\beta_{x_1}, ..., \beta_{x_{na}})'$,

[2]Bell (1984) points out that the problem with using polynomials in forecasting time series is the following. The assumption that the error terms are uncorrelated over time is virtually always unrealistic. It has the following effects. 1.) The behavior of long run forecasts is unreasonable (tending to $+\infty$ or -∞). 2.) If the fit of the curve at the end of the series is poor, short run forecasts are likely to be bad. 3.) Variances of forecast errors are usually highly unrealistic. ARIMA models tend not to suffer from these drawbacks.

[3]The central death rate of an individual with age x at time t is defined as a weighted average of the number of deaths in periods $t-1, t, t+1$, divided by a weighted average of the exposure for individuals with age $x-2, ..., x+2$. For details, see Benjamin and Pollard (1993).

$\{\gamma_t\}_{t=1}^T$ satisfy (2.8), then for any scalar c, $\alpha - \beta c$, β, $\gamma_t + c$, or α, $c\beta$, $\frac{1}{c}\gamma_t$ also satisfy (2.8). Therefore, Lee and Carter (1992) normalize by setting the sum of β_x to unity, $\sum_{x=1}^{na} \beta_x = 1$, and by imposing the constraint $\sum_{t=1}^{T} \gamma_t = 0$, implying that α_x becomes the population average over time of the age-specific log mortality rate $m_{x,t}$. The model in (2.8) then can be rewritten in terms of the mean centered log mortality rates $\widetilde{m}_t = (\widetilde{m}_{1,t}, ..., \widetilde{m}_{na,t})'$ as

$$\widetilde{m}_t = m_t - \alpha = \beta\gamma_t + \delta_t, \tag{2.9}$$

where $m_t = (m_{1,t}, ..., m_{na,t})'$, $\alpha = (\alpha_{x_1}, ..., \alpha_{x_{na}})'$, $\beta = (\beta_{x_1}, ..., \beta_{x_{na}})'$, and $\delta_t = (\delta_{x,t}, ..., \delta_{x,t})'$. Assuming a diagonal covariance matrix for δ_t, Lee and Carter (1992) propose to estimate the parameters via singular value decomposition (SVD). On the basis of a spectral decomposition of the covariance matrix $\frac{1}{T}X'X = V\Lambda V'$, with $X = (\widetilde{m}_1, ..., \widetilde{m}_T)'$ of mean centered age profiles, the matrix K of the principal components is given by $K = XV$, and the first column of K yields $\{\widehat{\gamma}_t\}_{t=1}^T$ with a zero mean[4]. Subsequently, each $\widehat{\beta}_x$ can be found by regressing, without a constant term, $m_{x,t} - \widehat{\alpha}_x$ on $\widehat{\gamma}_t$, separately for each age group x.[5]

Lee and Carter (1992) suggest a "second stage estimation", because the SVD method produces, in general, discrepancies between the estimated and the actual mortality rates, due to the fact that the model fits the log mortality rates instead of the mortality rates. This bias is removed by finding an adjusted mortality index $\{\widetilde{\gamma}_t\}_{t=1}^T$, which equates the model-implied death numbers to the observed ones in each year t:

$$\sum_x D_{x,t} = \sum_x E_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \widetilde{\gamma}_t), \forall t, \tag{2.10}$$

where $E_{x,t}$ and $D_{x,t}$ are the exposure to risk[6] and the actual number of death at age x and time t respectively. The $\{\widetilde{\gamma}_t\}_{t=1}^T$ satisfying (2.10) can be determined by an iterative procedure.

Finally, the Box-Jenkins approach is applied in order to find an appropriate ARIMA time-series model for the mortality index $\{\widetilde{\gamma}_t\}_{t=1}^T$.

Lee and Carter (1992) calculate τ-period ahead projections $\widehat{m}_{T+\tau}$ starting at T as

[4]If the i-th columns of X and K are denoted by $\mathbf{x}_i$ and $\mathbf{k}_i$, respectively, and v_{ij} denotes the $< i, j >$-th component of the matrix V, then $\mathbf{k}_1 = \sum_{i=1}^{na} v_{i1}\mathbf{x}_i$. Since the $\mathbf{x}_i$-s are mean centered log mortality rates, $\mathbf{k}_1$ also has mean zero.

[5]The normalization for β is achieved by scaling the estimate for β and γ_t by a constant $c = \sum_{x=1}^{na} \widehat{\beta}_x$ such that $\widehat{\beta}$ is replaced by $\frac{\widehat{\beta}}{c}$, and $\widehat{\gamma}_t$ is replaced by $c\widehat{\gamma}_t$.

[6]The exposure is the number of person years with age x in year t. For more details, see Gerber (1997).

follows:

$$\widehat{m}_{T+\tau} = \widehat{\alpha} + \widehat{\beta}\widetilde{\gamma}_{T+\tau}, \tag{2.11}$$

where $\widetilde{\gamma}_{T+\tau}$ is the τ-period ahead forecast of the latent process. Forecast errors, including parameter risk can be calculated based on bootstrapping the joint distribution of the estimated model parameters.

The estimation procedure suggested by Lee and Carter (1992) uses singular value decomposition which assumes homoskedasticity of errors over all ages, which might not always hold (Lee and Miller, 2001; Brouhns et al., 2002). Several alternative estimation approaches were proposed. Wilmoth (1993) applied the weighted least squares method (WLS), where the residuals were weighted by the number of deaths for every age group in each time period and the solutions of the parameters were found by an iterative procedure. Brouhns et al. (2002) implement the Lee-Carter model in a Poisson error setting. Instead of modeling the log of the mortality rates, they model the integer-valued number of deaths as a Poisson distributed random variable. Brouhns et al. (2002) considered

$$D_{x,t} \sim Poisson(E_{x,t}\mu_{x,t}) \text{ with } \mu_{x,t} = exp(\alpha_x + \beta_x\gamma_t), \tag{2.12}$$

where the meaning of the parameters is the same and also subject to similar normalization constraints as in the Lee and Carter (1992)-model.

Instead of applying the SVD to estimate α_x, β_x, and γ_t, Brouhns et al. (2002) determined these parameters by maximizing the log-likelihood of the model

$$L(\alpha, \beta, \gamma) = \sum_{x,t}\{D_{x,t}(\alpha_x + \beta_x\gamma_t) - E_{x,t}exp(\alpha_x + \beta_x\gamma_t)\} + \text{constant}. \tag{2.13}$$

Because of the presence of the bilinear term $\beta_x\gamma_t$, an iterative algorithm is used which solves the likelihood equations. Brouhns et al. (2002) claim that there is no need of a "second stage estimation" of $\widehat{\gamma}_t$ to equate the model-implied death numbers to the observed ones, because the observed number of deaths is modeled directly in the Poisson regression approach, instead of the transformed mortality rates in Lee and Carter (1992)-model

The Box-Jenkins methodology is used to find the appropriate ARIMA model for the estimated latent process $\{\widehat{\gamma}_t\}_{t=1}^{T}$, and future projections can be implemented similarly to the method proposed by Lee and Carter (1992).

Girosi and King (2005b) proposed a reformulation of the empirically quite often found version of the Lee and Carter (1992)-model, namely the version with a single latent factor, resulting in a random walk with drift. This version of the Lee and Carter model is given by :

$$m_t = \alpha + \beta\gamma_t + \delta_t, \tag{2.14}$$

with $\{\gamma_t\}_{t=1}^{T}$ following a random walk with drift c

$$\gamma_t = \gamma_{t-1} + c + \epsilon_t, \tag{2.15}$$

where ϵ_t represents the innovation term.

Following Girosi and King (2005b) we can rewrite this version of the Lee and Carter (1992)-model in (2.14) and (2.15), yielding

$$\begin{aligned} m_t &= \alpha + \beta\gamma_t + \delta_t &(2.16)\\ &= \beta c + (\alpha + \beta\gamma_{t-1} + \delta_{t-1}) + (\beta\epsilon_t + \delta_t - \delta_{t-1}) &(2.17)\\ &= \theta + m_{t-1} + \zeta_t &(2.18) \end{aligned}$$

with

$$\theta = \beta c,\ \zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}. \tag{2.19}$$

In the random walk with drift reformulation in (2.18) proposed by Girosi and King (2005b), the drift vector $\theta = (\theta_1, ..., \theta_{na})'$ and the covariance matrix $\Sigma_{\zeta|GK} \in \mathbb{R}^{na\times na}$ of ζ_t are arbitrary and not subject to any structure, and the error terms ζ_t could be either correlated or uncorrelated over time. In this reformulation the log central death rates (or some other way to measure log mortalities) are directly modeled as random walks with drift, making estimation and forecasting rather straightforward, simplifying considerably the original Lee and Carter estimation and prediction approach. Indeed, with $\Delta m_t = m_t - m_{t-1}$, we can estimate θ simply by the time average of Δm_t, i.e., by

$$\widehat{\theta}_T = \frac{1}{T-1}\sum_{t=2}^{T}\Delta m_t = \frac{1}{T-1}\left(m_T - m_1\right). \tag{2.20}$$

This estimator has well-known (T-asymptotic) characteristics. Predictions of future values of $m_{T+\tau}$, for $\tau = 1, 2, ...$, as well as the construction of the corresponding prediction intervals, can be based upon

$$m_{T+\tau} = m_T + \theta\tau + \sum_{t=T+1}^{T+\tau}\zeta_t. \tag{2.21}$$

For instance, Girosi and King (2005b), ignoring the moving average character of the error terms ζ_t, construct as predictors of $m_{T+\tau}$

$$\widehat{E}\left(m_{T+\tau}\,|\mathcal{F}_T\right) = m_T + \widehat{\theta}_T\tau. \tag{2.22}$$

Thus, as prediction for a particular age(-group) x, one can simply take the straight line going through the corresponding components of m_1 and m_T, extrapolated into the future.

The Girosi and King (2005b) random walk with drift formulation in (2.18) is equivalent with the Lee and Carter (1992)-model in (2.14) that is driven by a random walk with drift latent process in (2.15) if the structure in (2.19) preserved. Adding the Lee-Carter normalization $\sum_{x=1}^{na} \beta_x = 1$ yields $c = \omega$ and $\beta = \frac{\theta}{\omega}$, where $\omega = \sum_{x=1}^{na} \theta_x$. Then we can rewrite (2.18) as

$$m_t = \theta + m_{t-1} + \left(\frac{1}{\omega} \theta \epsilon_t + \delta_t - \delta_{t-1} \right). \tag{2.23}$$

Therefore, the covariance matrix $\Sigma_{\zeta|LC}$ of the noise ζ_t in the random walk with drift model that is equivalent with the Lee and Carter (1992) specification becomes

$$\Sigma_{\zeta|LC} = \sigma_\epsilon^2 \frac{1}{\omega^2} \theta \theta' + 2\Sigma_\delta. \tag{2.24}$$

This shows that in the Lee-Carter model shocks to mortality can be of two kinds. The term $\delta_t - \delta_{t-1}$ with variance $2\Sigma_\delta$ describes shocks that are uncorrelated across age groups, following from the assumption of Lee and Carter (1992). The term $\frac{1}{\omega}\theta\epsilon_t$ with variance $\sigma_\epsilon^2 \frac{1}{\omega^2}\theta\theta'$ describes shocks that are perfectly correlated across age groups, and the size of the perfectly correlated shocks is restricted to be $\beta\epsilon_t$. It implies that age group x with higher sensitivity β_x to the underlying latent process γ_t, that has been declining faster than others, receives larger shocks. By keeping the structure of the Lee-Carter specification, Girosi and King (2005b) claim that shocks to mortality other than those that are perfectly correlated or uncorrelated across age groups will be missed by the model.

The main difference between the general random walk with drift and the Lee-Carter specification lies in the nature of the shocks to mortality. In the Lee-Carter model the error term ζ_t is restricted in a way which explicitly depends on the drift vector θ, and ζ_t is autocorrelated with a first-order moving average structure $\delta_t - \delta_{t-1}$, while $\Sigma_{\zeta|GK}$ is arbitrary with no structure for the autocorrelation in ζ_t.

Girosi and King (2005b) showed that if the data are generated according to the Lee and Carter (1992)-model, then the estimate for the drift parameter θ in the random walk with drift model with arbitrary drift vector and covariance matrix is unbiased. However, Girosi and King (2005b) concluded that this estimate for θ is less efficient, than the one obtained by the Lee and Carter (1992)-method. If the data are generated according to the more general random walk with drift model with arbitrary drift and covariance matrix, then the drift parameter estimated by the Lee and Carter (1992)-method will be biased. It implies that the Lee-Carter estimator, and therefore keeping the structure suggested in (2.19) is preferable to the random walk with drift reformulation with an arbitrary covariance matrix only when the modeler has high confidence in its underlying assumptions.

To deal with the potential moving average character[7] of the error term ζ_t, one could maintain the structure $\zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}$ following from Lee and Carter (1992), or, alternatively, one could postulate that ζ_t follows an MA(1)-structure given by

$$\zeta_t = \xi_t + \Theta\xi_{t-1}, \tag{2.25}$$

with Θ an $(na \times na)$-matrix of unknown parameters, and where ξ_t is an na-dimensional vector of white noise with an arbitrary covariance matrix $\Sigma_\xi \in \mathbb{R}^{na\times na}$, satisfying the distributional assumptions

$$\xi_t \,|\mathcal{F}_{t-1} \sim (0, \Sigma_\xi)\,.$$

With these modifications, the Lee and Carter (1992)-model becomes

$$\begin{aligned} m_t &= \theta + m_{t-1} + \zeta_t, && (2.26)\\ \zeta_t &= \xi_t + \Theta\xi_{t-1}, && (2.27)\\ \xi_t \,|\mathcal{F}_{t-1} &\sim (0, \Sigma_\xi)\,. \end{aligned}$$

This reformulation maintains the arbitrary structure of the covariance matrix as it was proposed by Girosi and King (2005b), and it takes into account the potential autocorrelation between the error terms ζ_t.

Koissi and Shapiro (2006) proposed a fuzzy formulation of the Lee-Carter model. The authors use a fuzzy logic estimation approach, where the errors are viewed as fuzziness of the model structure, and the potential heteroskedasticity is not an issue.

The original single-factor model suggested by Lee and Carter seems to be too rigid to describe the historical evolution of death rates. Chapter 1 already indicated that the mortality of the young, adult, and elderly population is likely driven by factors with different properties. A single factor is not able to reproduce the cross-sectional variation in the age-specific mortality rates. The mortality of some groups is reproduced with a better fit, while for some other groups, the fit of the model is relatively poor.

[7] Specification tests indicate (see Chapter 3, for instance) that the random walk with drift reformulation of the Lee-Carter model by Girosi and King (2005b) (with arbitrary drift and covariance matrix and uncorrelated error terms in time) estimated for the Dutch mortality data violates the no autocorrelation assumption.

It is also reflected in the systematic error structure of the model. Therefore, Lee and Miller (2001), Carter and Prskawetz (2001), and Booth et al. (2002) suggest that time variation in the parameters is necessary to fit the data adequately. As an alternative solution, an "age-specific enhancement" of the Lee-Carter model is considered by including the second unobserved latent factor in Renshaw and Haberman (2003a), where both factors capture the period effects. Renshaw and Haberman (2003a) find that the in-sample fit of the extended model improves, and the model structure is sufficiently flexible to represent adequately all the age-specific differences, and no time variation in the age-specific parameters is necessary. Renshaw and Haberman (2006) proposed an alternative extension of the original single-factor Lee-Carter methodology by adding age-specific cohort effects to the existing age-specific period effects. The period effects are captured by the time-varying latent factor through the age-specific factor loadings as it was suggested by Lee and Carter (1992), and, in addition, Renshaw and Haberman introduce an additional factor which is varying with the year of birth of the cohorts, suggesting that birth cohorts have common characteristics which are present over the lifetime of a certain cohort.

Lee (2000) suggests to use the Lee-Carter methodology for the extrapolation of mortality trends by mortality reduction factors, while a Poisson-based equivalent approach was proposed by Renshaw and Haberman (2003b).

Another recent strand of the literature models the mortality with postulating typically a mean reverting process. The force of mortality has an exponentially affine structure, so that the results of the term structure of interest rates literature can be applied. For instance Milevsky and Promislow (2001) model the force of mortality equivalent to a Gompertz model with a mean-reverting, time-varying scaling factor. Dahl (2004) and Biffis (2005) also model the force of mortality as a stochastic affine class process. Schrager (2006) also proposed an affine stochastic mortality model with an underlying multifactor latent process which follows a mean-reverting square-root diffusion. Cairns et al. (2006b) propose a mortality model where the realized 1-year mortality rates are driven by 2-factor Perks stochastic processes.

2.3 Longevity risk in mortality projections

The primary objective of mortality modeling is to produce out-of-the-sample forecasts of mortality. Projecting mortality for age groups with few or no observations is important (for example, projecting mortality for the old, see Lindbergson, 2001; Coale and Guo, 1989; Coale and Kisker, 1990). However, these are not the type of applications we have in mind. In this section we focus on the literature on the uncertainty in mortality forecasts, its effect on the price of mortality related financial products, and solvency positions of institutions selling these products.

Longevity risk is related to the fact that the remaining lifetime of an individual is uncertain. The uncertainty which contributes to the total risk can be decomposed into several components. We distinguish *micro-longevity risk*, which results from nonsystematic deviations from an individual's expected remaining lifetime, and *macro-longevity risk*, which results from the fact that survival probabilities change over time related to the uncertainty in the stochastic latent process driving the mortality evolutions. Moreover, additional sources of risk are the *parameter risk* related to the estimation risk of the model parameters given a model, and *model risk* capturing the risk in an inappropriate model specification.

The studies of Olivieri (2001, 2002), Coppola et al. (2000, 2003a,b), Di Lorenzo and Sibillo (2002), Pitacco (2002) look at the effect of macro- and micro-longevity risk on the riskiness of a pension annuity contract. Similarly, Olivieri and Pitacco (2003) calibrate solvency buffers for life annuity portfolios related to longevity risk. They find that the micro-longevity risk for an annuity portfolio (measured by the variance of the payoff) becomes unimportant when the size of the portfolio becomes large. In contrast, the size of macro-longevity risk is independent of portfolio size. The results of the studies clearly raise the issue on the importance of longevity risk in mortality projections.

In the applications of the Lee and Carter (1992) method the only risk source, that we call macro-longevity risk (the only source of uncertainty is due to the stochastic latent process), was quantified. Lee (2000) claims that the uncertainty produced solely by the macro-longevity risk is too narrow and it understates the uncertainty about the future level of life expectancy, since it does not take into account uncertainty arising from errors in the estimations, nor the uncertainty about the model specification. Brouhns et al. (2005) used bootstrapping method for the Poisson maximum likelihood method to quantify the risk in the estimated parameters. Koissi et al. (2006) estimated the Lee-Carter model with different estimation techniques (singular value decomposition, maximum likelihood and weighted least squares methods) for Finland and used the maximum likelihood estimation method to forecast future expected remaining lifetime with parameter risk in the Finnish population. Dowd et al. (2006) apply a recently developed 2-factor mortality model of Cairns et al. (2006a) to estimate risk measures for mortality-dependent positions, such as a long position in zero-coupon or coupon-paying longevity bonds, or a combination of a short and long positions in mortality dependent coupon bonds with no or some basis risk left due to different reference populations. The analysis considers both macro-longevity and parameter risk. Their results suggest that mortality-dependent positions can be very risky. While in money's worth terms, part of the risk arising in a long horizon is amortized by the discounting effect, in relative terms (measuring the risk relative to the money's worth of the position), there is often a considerable amount of risk at the end of the maturity spectrum.

Cossette et al. (2005) estimate the model suggested by Brouhns et al. (2002) for a population and use a relational model embedded in a Poisson regression approach to create the mortality tables of a given pension plan by using the population mortality characteristics. The paper looks at the effect of mortality improvement on the expected remaining lifetime, annuity prices, and solvency of pension plans, where the benchmark was a static period mortality table.

Khalaf-Allah et al. (2006) use a deterministic trend model of Sithole et al. (2000) and measures the effect of mortality improvement on the cost of annuities in the UK. The paper also considers the effect of parameter uncertainty on the projected distribution of the annuity cost. The expected present value of annuity without mortality improvement is compared to the case when mortality improvement is allowed. For a flat yield curve at 6%, the improvement in mortality had an effect of 3% increase in the expected annuity value for 65-year-old men and 6% increase for women.

2.4 Heterogeneity in survival rates

In Chapter 1 we already presented the evidence on the heterogeneity in human survival rates by using the data of the Human Mortality Database and the findings of several publications (for instance, Huisman et al., 2004, 2005). We will not reproduce them in this section once more, instead, we focus on the pricing implications of the survival heterogeneity.

Kwon and Jones (2006) explore the heterogeneity of survival rates in the Canadian population which is related to risk factors such as socioeconomic/demographic status (age, sex, education, income, occupation or marital status) behavioral risk factors (smoking, alcohol intake, etc.) and health indicators (blood pressure, cholesterol level, etc.). Kwon and Jones (2006) estimated a discrete time multiple state Markov chain model with age-specific transition matrices, which allow for the variability of specific risk factors over time. The results of the analysis imply that the present value of a life annuity and the single premium of term insurance contracts are more favorable for women, for married, for people with higher income, or for the ones who do not smoke, etc., which suggests that the heterogeneity in survival rates, and, therefore, the risk factors should be reflected in pricing, so that annuity holders or insureds pay fair values for life insurance or annuity products.

Brown (2002, 2003) also documented heterogeneity in survival rates among cohorts grouped along socioeconomic, ethnic or racial lines, and its effect on the money's worth of participation in a compulsory annuitization framework in the US. The money‘s worth measure is the expected present value of annuity payments per money amount spent to purchase the annuity. Brown (2002, 2003) report the money‘s worth of the uniform annuities for individuals taking into account cohort-specific (gender, educational, race) survival characteristics. Brown finds that the money's worth of participation of cohorts with lower than average survival prospects is less than for the ones with higher than average survival rates. It clearly implies a wealth redistribution among cohorts due to the uniform pricing which ignores group-specific survival differences.

Feldstein and Liebman (2002) calculated the net present value of the lifetime participation for different cohorts in the US population in a funded pension system. Annuities at retirement are calculated by using a single uniform unisex mortality table, disregarding individual survival characteristics. The results are similar to Brown (2002, 2003), because wealth is redistributed from men to women, from black to white, and from low educated to higher educated.

2.5 Contribution of the thesis

Chapter 3 introduces a model for human mortality rates. In modeling and forecasting mortality the Lee-Carter approach (Lee and Carter, 1992) is the benchmark methodology. In many empirical applications the Lee-Carter approach results in a model that describes the log central death rate by means of a linear trend, where different age groups have different trends. However, due to the volatility in mortality data, the estimation of these trends, and, thus, the forecasts based on them, are rather sensitive to the sample period employed. We allow for time-varying trends, depending on a few underlying factors, to make the estimates of the future trends less sensitive to the sampling period. We formulate our model in a state-space framework, and use the Kalman filtering technique to estimate it. We illustrate our model using Dutch mortality data.

Chapter 4 analyzes the importance of longevity risk for the solvency of a portfolio of pension annuities. We use the generalized 2-factor Lee-Carter mortality model introduced in Chapter 3 to produce forecasts of future mortality rates, and to assess the relative importance of micro- and macro-longevity risk and parameter risk for funding ratio uncertainty. The results show that if uncertainty in future lifetime is the only source of uncertainty (and future mortality improvement was taken into account when expected liabilities are calculated, interest and investment risk were assumed to be fully diversified) pension funds are exposed to a substantial amount of risk. For large portfolios, systematic deviations from expected survival probabilities and parameter risk imply that buffers that reduce the probability of underfunding to 2.5% at a 5-year horizon have to be of the order of magnitude of 7.1% of the value of the initial liabilities. Alternatively, longevity risk could be hedged by means of stop loss reinsurance contracts. We use the mortality forecast model to price these contracts. The relative size of mortality risk becomes less important in the total risk of pension funds, if the assets are exposed to a substantial amount of investment risk.

Chapter 5 measures the present value of a single year participation in a collective scheme consisting of heterogeneous participants. In many countries, employees have implicit or explicit options to opt out of collective pension schemes. The contributions to such schemes are often set uniformly, irrespective of age, gender, or education level. We quantify the incentives for individuals that participate in such systems. We show for instance, that young males with low education have a strong incentive to opt out of the collective system in the case of uniform pricing, since their contribution is high relative to the benefit to be obtained. This incentive is enforced by the fact that the switching costs for young individuals are relatively low. Moreover, it turns out that the indexation quality of the scheme is a non-negligible determinant of the present value of participation, and it introduces additional incentives to opt out of schemes with inadequate funding.

Chapter 3

Estimating the Term Structure of Mortality

3.1 Introduction

For life-related insurance products, one can distinguish two types of actuarial risk. First, institutions offering products depending on the lifetime of an individual face risk, simply because lifetime is uncertain. However, it is well known that this type of risk reduces significantly when the portfolio size is increased. Second, mortality patterns may change over time due to, for example, improvements in the standards of living and lifestyle or better prospects in the medical system. This source of risk can clearly not be diversified away by increasing the portfolio size. As a consequence, changes in survival probabilities can have a major effect on, for example, fair premiums for life insurance or funding ratios for pension funds. Therefore, forecasting future mortality risk is in the interest of insurance companies and pension funds.

Several methods for capturing the behavior of mortality rates over time and for forecasting future mortality have been developed. The literature evolved along several directions. The *deterministic trend* approach fits curves as a function of age and time to approximate mortality rates. Fitting curves to mortality rates goes back to Gompertz or Makeham in the 19th century. These early efforts tried to fit part of the mortality curve by considering only the age dimension, typically the middle and elderly aged. Heligman and Pollard (1980) already fitted a curve to the entire age range, but they did not estimate the time effect either. Most recent models fit curves to mortality rates in both the age and time dimension. For instance, Renshaw et al. (1996) use polynomials to describe the age and time evolution of mortality changes. These models give a very accurate in-sample fit. However, a main disadvantage of this deterministic trend approach is that the accurate in-sample fit is translated into quite small prediction intervals, when extrapolated out of sample, but such accurate predictions do not seem to be very realistic, also because of the model uncertainty that is usually not taken into account.

The *stochastic trend* methodology seems to be a more parsimonious approach, which tries to explain the variability of mortality rates with a low number of unobserved latent factors: death rates are explained as a function of time-varying unobserved state variables and age-specific parameters, which describe the relative sensitivities of individual age groups to the change in the underlying unobserved state variables. The stochastic trend approach was first introduced for mortality forecasts in Lee and Carter (1992). They explore the time-series behavior of mortality movements between age groups by using a single latent factor, which is responsible for describing the general level of log mortality. Log central death rates are modeled as the sum of a time-invariant, age-specific constant, and the product of an age-specific time invariant component and the time-varying latent factor. The age-specific component represents the sensitivity of an individual age group to the general level of mortality changes. The estimation of the model proceeds in several steps. First, singular value decomposition (SVD) is used to retrieve the underlying factor. Second, the age specific parameters are estimated by means of ordinary least squares. Then the latent factor is re-estimated while keeping age-specific parameters from the first step constant, in order to guarantee that the sum of the implied number of deaths equals the sum of the actual number of deaths in each time period. Finally, ARIMA modeling is used to fit a time series process to the latent variable, which can be used to make forecasts. In case of Lee and Carter (1992) the time process of the latent factor turned out to be a random walk with drift, implying that its forecast is just a linear trend, but with a prediction interval much wider than obtained in case of a deterministic trend approach.

A whole strand of literature evolved from the original Lee-Carter approach, see, for example, Lee and Miller (2001), Carter and Prskawetz (2001), Booth et al. (2002), Brouhns et al. (2002), and Renshaw and Haberman (2003a,b) to mention just a few. Recently, Girosi and King (2005b) proposed a reformulation of the empirically quite often found version of the Lee and Carter (1992)-model, which is the version having a single latent factor, following a random walk with drift. In this reformulation the log central death rates (or some other way to measure log mortalities) are directly modeled as random walks with drift, making estimation and forecasting a rather straightforward exercise in econometrics, simplifying considerably the original Lee and Carter (1992)-estimation and prediction approach.

We take this reformulation by Girosi and King (2005b) as our starting point. When using actual mortality data to estimate this version of the Lee and Carter (1992)-model, we make the following observation. First, the typical sample period is rather short, usually starting somewhere in the nineteenth century, resulting in only around 150 annual observations (or even less). Secondly, the observed mortality data turns out to be quite volatile, particularly, during the nineteenth century, but also around, for instance, the first and second world war. This implies that the estimation of the drift term in the Girosi and King (2005b) reformulation – the slope of the long run trend – might be rather sensitive to the sample period used in estimation, making also the long run forecasts sensitive to the sample period.

To account for this sensitivity, we propose to extend the Girosi and King (2005b) formulation of the Lee and Carter (1992)-model by making the drift term time dependent. We postulate that this time dependent drift term is a (time-independent) affine transformation of a few underlying (time-varying) latent factors, which capture the time movements, common to all different age groups. The underlying latent factors are assumed to have a long-run zero mean, but their short run sample means might deviate from zero. These non-zero sample means could be used to extract a long run trend that might be less sensitive to the sample period employed.

The model is set up in a state-space framework, well-known from time series modeling. This makes the use of the Kalman filtering technique possible, still allowing econometric estimation and prediction in a rather straightforward way, as in the Girosi and King (2005b) reformulation of the Lee and Carter (1992)-model.

The remainder of the chapter is organized as follows. In Section 3.2, we first provide a description of the Lee-Carter model, including the reformulation by Girosi and King (2005b), and discuss some of the drawbacks of this way of modeling mortality. Section 3.3 introduces our approach, which we illustrate in Section 3.4 using Dutch data on mortality. Section 3.5 concludes.

3.2 The Lee and Carter approach

In this section we first describe the Lee and Carter (1992)-approach and the way it is usually estimated. Then we present the reformulation presented in Girosi and King (2005b) and we present the forecasting of future mortality based on the reformulation by Girosi and King (2005b), together with some of the limitations of this way of modeling and forecasting mortality.[1] In the next section we then introduce our alternative approach.

Let D_{xt} be the number of people with age x that died in year t, and E_{xt}, the exposure being the number of person years[2] with age x in year t, with $x \in \{1, ..., na\}$, and $t \in \{1, ..., T\}$. We consider modelling of[3]

$$m_{xt} = \ln\left(\frac{D_{xt}}{E_{xt}}\right). \tag{3.1}$$

Define

$$m_t = \begin{pmatrix} m_{1,t} \\ \vdots \\ m_{na,t} \end{pmatrix}, \tag{3.2}$$

then the model according to Lee and Carter (1992) can be formulated as

$$m_t = \alpha + \beta\gamma_t + \delta_t, \tag{3.3}$$

with unknown parameter vectors $\alpha = (\alpha_1, ..., \alpha_{na})'$ and $\beta = (\beta_1, ..., \beta_{na})'$, and a vector of (measurement) error terms $\delta_t = (\delta_{1,t}, ..., \delta_{na,t})'$, where

$$\{\gamma_t\}_{t=1}^{T}$$

is a one-dimensional underlying latent process, assumed to be governed by

$$\gamma_t = c_0 + c_1\gamma_{t-1} + ... + c_k\gamma_{t-k} + \epsilon_t, \tag{3.4}$$

with unknown parameters c_0, c_1,..., c_k, and error term ϵ_t satisfying

$$\epsilon_t = \omega_t + d_1\omega_{t-1} + ... + d_\ell\omega_{t-\ell}, \tag{3.5}$$

with unknown parameters d_0, d_1,..., d_ℓ, where the error term ω_t and error term vector δ_t are white noise, satisfying the distributional assumption

$$\begin{pmatrix} \delta_t \\ \omega_t \end{pmatrix} |\mathcal{F}_{t-1} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\delta & 0 \\ 0 & \sigma_\omega^2 \end{pmatrix}\right),$$

[1]For a detailed exposition of the link between the Lee and Carter (1992)-model and the Girosi and King (2005b)-reformulation, see Chapter 2.

[2]For more details on the definition and the estimation of E_{xt}, see Gerber (1997)

[3]Lee and Carter (1992) use the log of the *central death rate*. The central death rate of an individual with age x at time t is defined as a weighted average of the number of deaths in periods $t-1, t, t+1$, divided by a weighted average of the exposure for individuals with age $x-2, ..., x+2$. For additional details, see Benjamin and Pollard (1993).

with $\mathcal{F}_{t-1}$ representing the information up to time $t-1$, and with Σ_δ the unknown covariance matrix of δ_t and σ^2_ω the unknown variance of ω_t. The error term ω_t driving the γ_t-process is assumed to be uncorrelated with the vector of error terms δ_t appearing in the m_t-equation.

As originally proposed by Lee and Carter (1992), the model is usually estimated in several steps. In the first step, Singular Value Decomposition (SVD) is applied to retrieve the underlying latent process, yielding $\{\widehat{\gamma}_t\}_{t=1}^T$. Secondly, OLS regressions are run for each age group $x = 1, ..., na$, to estimate the age-specific parameters, resulting in $\widehat{\alpha}$ and $\widehat{\beta}$. Thirdly, the estimated $\{\widehat{\gamma}_t\}_{t=1}^T$ is adjusted to ensure equality between the observed and model-implied number of deaths in a certain period, i.e., $\{\widehat{\gamma}_t\}_{t=1}^T$ is replaced by $\{\widetilde{\gamma}_t\}_{t=1}^T$ such that:

$$\sum_{x=1}^{na} D_{xt} = \sum_{x=1}^{na} \left[E_{xt} \exp(\widehat{\alpha}_x + \widehat{\beta}_x \widetilde{\gamma}_t) \right]. \tag{3.6}$$

Finally, the Box-Jenkins method is used to identify and estimate the dynamics of the latent factor $\widetilde{\gamma}_t$.[4]

Typically, when estimating the Lee and Carter (1992)-model, one usually infers that

$$c_0 = c,\ c_1 = 1,\ c_2 = c_3 = ... = 0,\ d_1 = d_2 = ... = 0,$$

meaning that the underlying latent process is a random walk with drift. Thus, the typical version of the Lee and Carter (1992)-model, that is estimated and applied in forecasting, is given by

$$m_t = \alpha + \beta\gamma_t + \delta_t, \tag{3.7}$$

with

$$\gamma_t = c + \gamma_{t-1} + \epsilon_t, \tag{3.8}$$

where

$$\begin{pmatrix} \delta_t \\ \epsilon_t \end{pmatrix} |\mathcal{F}_{t-1} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\delta & 0 \\ 0 & \sigma^2_\epsilon \end{pmatrix} \right).$$

[4]The readjustment of the latent process in the third step is done in order to avoid sizeable differences between the observed and the model-implied number of deaths. Other advantages of the readjustment have been mentioned in Lee (2000). However, the fact that the readjustment is done without re-estimating the age-specific sensitivity parameters $\widehat{\beta}_x$ also has several drawbacks. First, since the estimated variables $\widehat{\gamma}_t$, obtained in the first step, are adjusted after the age-specific coefficients in the OLS regressions are estimated, the resulting term $\widehat{\beta}_x\widetilde{\gamma}_t$ might not accurately describe the movements in the log death rates $m_{x,t}$ anymore. Second, the standard error estimated for $\widehat{\beta}_x$ in the age-specific regressions does not necessarily describe the correct size of the uncertainty for the parameter estimates.

Following Girosi and King (2005b) we can rewrite this version of the Lee and Carter (1992)-model, yielding

$$\begin{aligned} m_t &= \alpha + \beta\gamma_t + \delta_t && (3.9) \\ &= \beta c + \alpha + \beta\gamma_{t-1} + \delta_{t-1} + (\beta\epsilon_t + \delta_t - \delta_{t-1}) && (3.10) \\ &= \theta + m_{t-1} + \zeta_t && (3.11) \end{aligned}$$

with

$$\theta = \beta c,\ \zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}.$$

As noted by Girosi and King (2005b), the typical Lee and Carter (1992)-model rewritten in this way, can easily be estimated and predicted. Indeed, with $\Delta m_t = m_t - m_{t-1}$, we can estimate θ simply by the time average of Δm_t, i.e., by

$$\widehat{\theta}_T = \frac{1}{T-1}\sum_{t=2}^{T}\Delta m_t = \frac{1}{T-1}\left(m_T - m_1\right). \tag{3.12}$$

This estimator has well-known (T-asymptotic) characteristics. Predictions of future values of $m_{T+\tau}$, for $\tau = 1, 2, ...$, as well as the construction of the corresponding prediction intervals, can be based upon

$$m_{T+\tau} = m_T + \theta\tau + \sum_{t=T+1}^{T+\tau}\zeta_t. \tag{3.13}$$

For instance, Girosi and King (2005b), ignoring the moving average character of the error terms ζ_t, construct as predictors of $m_{T+\tau}$

$$\widehat{E}\left(m_{T+\tau}\left|\mathcal{F}_T\right.\right) = m_T + \widehat{\theta}_T\tau. \tag{3.14}$$

Thus, as prediction for a particular age(-group) $x \in \{1, ..., na\}$, one can simply take the straight line going through the corresponding components of m_1 and m_T, extrapolated into the future.

To deal with the potential moving average character of the error term ζ_t, one could maintain the structure $\zeta_t = \beta\epsilon_t + \delta_t - \delta_{t-1}$ following from Lee and Carter (1992), or, alternatively, one could postulate that ζ_t follows an MA(1)-structure given by

$$\zeta_t = \xi_t + \Theta\xi_{t-1}, \tag{3.15}$$

with Θ an $(na \times na)$-matrix of unknown parameters, and where ξ_t is an na-dimensional vector of white noise, satisfying the distributional assumptions

$$\xi_t\left|\mathcal{F}_{t-1}\right. \sim \left(0, \Sigma_\xi\right).$$

With these modifications, the Lee and Carter (1992)-model becomes

$$m_t = \theta + m_{t-1} + \zeta_t, \quad (3.16)$$

$$\zeta_t = \xi_t + \Theta \xi_{t-1}, \quad (3.17)$$

$$\xi_t \,|\mathcal{F}_{t-1} \sim (0, \Sigma_\xi)\,.$$

A main drawback of the Lee and Carter (1992)-model follows from the Girosi and King (2005b)-specification. Ignoring for simplicity the possible forecast correction due to an MA-error term (which only affects the level but not the slope), the forecast of age (-group) $x \in \{1, ..., na\}$ is essentially the straight line through $m_{x,1}$ and $m_{x,T}$, extrapolated into the future. Figure 3.1 shows Dutch mortality data of the age group 50-54 years during the sample period 1850 to 2003. As this figure illustrates, the mortality data is rather volatile, particularly at the beginning of the sample period, but also around the first and second world wars. This means that the estimates, and, thus, the mortality forecasts, might be rather sensitive to the exact sample period used in estimation: The straight lines through $m_{x,t}$ and $m_{x,\tau}$ may be different for different values of t or τ, resulting in quite different long run forecasts.

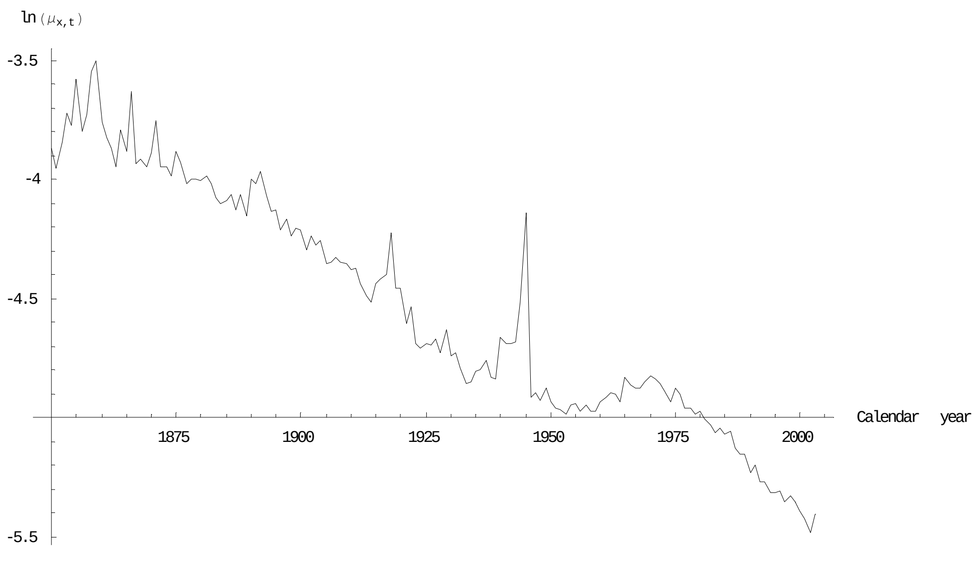

Figure 3.1: Log mortality for the age group of 50-54, men. The figure shows log mortality data of Dutch men for the age group of 50-54 years during the sample period 1850 to 2003. Data source: Human Mortality Database.

In the next section we present an extension of the Lee and Carter (1992)-model, starting from the Girosi and King (2005b)-reformulation, that is aimed to result in estimates of the long run trend that might be less sensitive to the particular sample period employed.

3.3 Lee-Carter with time-varying drift

In this section we present our generalization of the Lee and Carter (1992)-approach, taking as starting point the version typically found in empirical studies, as described in the previous section. We first describe the model and its motivation, then its estimation and its use in forecasting.

As generalization of the Lee and Carter (1992)-approach, we propose the following model for m_t:

$$m_t = \theta_t + m_{t-1} + \zeta_t \tag{3.18}$$

with

$$\begin{aligned} \theta_t &= a + Bu_t && (3.19)\\ u_t - \mu_u &= \Gamma\left(u_{t-1} - \mu_u\right) + \eta_t && (3.20) \end{aligned}$$

with u_t an nf-dimensional vector of underlying latent factors, driving the "constant" θ_t, where $a \in \mathbb{R}^{na}$, $B \in \mathbb{R}^{na\times nf}$, $\mu_u = E\left(u_t\right) \in \mathbb{R}^{nf}$, and $\Gamma \in \mathbb{R}^{nf\times nf}$ are unknown parameter vectors and matrices, where the na-dimensional vector of (measurement) errors ζ_t satisfies

$$\zeta_t = \xi_t + \Theta\xi_{t-1} \tag{3.21}$$

with $\Theta \in \mathbb{R}^{na\times na}$ a matrix with unknown parameters capturing the MA-effects, where the nf-dimensional vector of (measurement) errors η_t satisfies

$$\eta_t = \psi_t + \Xi\psi_{t-1} \tag{3.22}$$

with $\Xi \in \mathbb{R}^{nf\times nf}$ a matrix with unknown parameters capturing the MA-effects, and where the vectors of error terms ψ_t and ξ_t are white noise, satisfying the distributional assumption

$$\begin{pmatrix} \psi_t \\ \xi_t \end{pmatrix} |\mathcal{F}_{t-1} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\psi & 0 \\ 0 & \Sigma_\xi \end{pmatrix}\right)$$

with $\Sigma_\psi \in \mathbb{R}^{nf\times nf}$ and $\Sigma_\xi \in \mathbb{R}^{na\times na}$ the unknown covariance matrices of ψ_t and ξ_t, respectively.[5] The vectors of error terms ψ_t and ζ_t are assumed to be uncorrelated. The Lee and Carter (1992)-model is obtained as a special case by imposing $B = 0$.

[5] This model can easily be generalized to allow for a higher order moving average in the equation for

The long-run mean of the u_t-process is postulated to be equal to zero, i.e., $\mu_u = 0$, so that u_t corresponds to a process that fluctuates around zero. The changes in the drift term θ_t is postulated to be picked up by these changes in u_t. To capture comovements between different age groups, the time-varying short run changes are modeled as Bu_t, with u_t low dimensional and with B the age (-group) specific sensitivities to the underlying time-varying latent process.

To estimate the model, we apply the state-space method combined with the Kalman filtering technique (see, for instance, Durbin and Koopman, 2001; Hamilton, 1994). The model can straightforwardly be put in a state space form, with as 'observation equation'

$$\Delta m_t \equiv m_t - m_{t-1} = (a + B\mu_u) + \left[B \vdots 0 \vdots 0 \vdots I \vdots \Xi\right] \begin{pmatrix} u_t - \mu_u \\ \psi_t \\ \psi_{t-1} \\ \xi_t \\ \xi_{t-1} \end{pmatrix} \equiv A + H' z_t \quad (3.23)$$

and with 'state equation'

$$z_t = \begin{pmatrix} u_t - \mu_u \\ \psi_t \\ \psi_{t-1} \\ \xi_t \\ \xi_{t-1} \end{pmatrix} = \begin{pmatrix} \Gamma & \Xi & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{pmatrix} \begin{pmatrix} u_{t-1} - \mu_u \\ \psi_{t-1} \\ \psi_{t-2} \\ \xi_{t-1} \\ \xi_{t-2} \end{pmatrix} + \begin{pmatrix} \psi_t \\ \psi_t \\ 0 \\ \xi_t \\ 0 \end{pmatrix} \equiv F z_{t-1} + v_t. \quad (3.24)$$

Let Σ_v denote the covariance matrix of v_t. Following Hamilton (1994), we can easily derive the likelihood function corresponding to our state-space model specification, under the additional assumption that the distribution of

$$\begin{pmatrix} \psi_t \\ \xi_t \end{pmatrix} | \mathcal{F}_{t-1}$$

is normal. Let $\widehat{z}_{t|t-1}$ denote the best linear estimate of z_t, given information available up to time $t-1$, and let $P_{t|t-1}$ denote the forecast error, i.e.,

$$\widehat{z}_{t|t-1} = E\left(z_t \mid \mathcal{F}_{t-1}\right) = F z_{t-1} \quad (3.25)$$

$$P_{t|t-1} = E\left[\left(z_t - \widehat{z}_{t|t-1}\right)\left(z_t - \widehat{z}_{t|t-1}\right)' \mid \mathcal{F}_{t-1}\right] = \Sigma_v, \quad (3.26)$$

m_t or a higher order autoregression and higher order moving average in the equation for u_t, but for the sake of simplicity (and also based on empirical outcomes), we shall focus in the sequel on the current special version of the model. However, generalizing the subsequent analysis to the case with (higher order) moving average terms is straightforward.

where

$$\widehat{z}_{1|0} = E(z_1) = 0, \ P_{1|0} = E(z_1 z_1') = \Sigma_v.$$

Then, if the initial state and the innovations are multivariate Gaussian, the distribution of $\triangle m_t$, conditional on the information $\mathcal{F}_{t-1}$ up to time $t-1$ is normal:

$$\triangle m_t \mid \mathcal{F}_{t-1} \sim N\left(A + H'\widehat{z}_{t|t-1}, H'P_{t|t-1}H\right), \tag{3.27}$$

and the likelihood contribution for observation t is given by:

$$\begin{aligned} L_t &= (2\pi)^{-\frac{m}{2}} * \det\left(H'P_{t|t-1}H\right)^{-\frac{1}{2}} \times \\ &\exp\begin{pmatrix} -\frac{1}{2}\left(\triangle m_t - \left(A + H'\widehat{z}_{t|t-1}\right)\right)'\left(H'P_{t|t-1}H\right)^{-1} \\ \times\left(\triangle m_t - \left(A + H'\widehat{z}_{t|t-1}\right)\right) \end{pmatrix}. \end{aligned} \tag{3.28}$$

In the sequel we shall consider the likelihood as a quasi-likelihood, and calculate the asymptotic covariance matrix of the quasi-maximum likelihood accordingly, to allow for the possibility of nonnormally distributed error terms, thus, assuming only that the first moments are correctly specified.

The construction of predictions as well as the corresponding prediction intervals can be based upon

$$m_{T+\tau} = m_T + \sum_{t=T+1}^{T+\tau} (\theta_t + \zeta_t) \tag{3.29}$$

$$= m_T + \sum_{t=T+1}^{T+\tau} (a + Bu_t + \xi_t + \Theta\xi_{t-1}) \tag{3.30}$$

$$= m_T + a\tau + B\sum_{t=T+1}^{T+\tau} u_t + \sum_{t=T+1}^{T+\tau} (\xi_t + \Theta\xi_{t-1}). \tag{3.31}$$

As prediction of future values of $m_{t+\tau}$ we shall use

$$\widehat{E}(m_{T+\tau} | \mathcal{F}_T) = m_T + \widehat{a}\tau + \widehat{B}\sum_{t=T+1}^{T+\tau} \widehat{E}(u_t | \mathcal{F}_T) + \widehat{\Theta}\widehat{\xi}_T, \tag{3.32}$$

with

$$\widehat{E}(u_{T+t} | \mathcal{F}_T) = \left(\widehat{\mu}_u + \widehat{\Gamma}\widehat{\mu}_u + ... + \widehat{\Gamma}^{t-1}\widehat{\mu}_u\right) + \widehat{\Gamma}^t(\widehat{u}_T - \widehat{\mu}_u) + \widehat{\Gamma}^{t-1}\widehat{\Xi}\widehat{\psi}_T, \tag{3.33}$$

where the hats ($\hat{}$) indicate estimated values of the corresponding parameters.[6] Most estimators follow straightforwardly from maximizing the quasi-log-likelihood. Notice

[6]Prediction intervals can easily be constructed via simulation. Given the asymptotic distribution of the quasi-maximum likelihood estimator, we can simulate parameter values, and given these simulated parameter values, we can simulate the process for $m_{T+\tau}$ assuming that the white noise processes follow a normal distribution. In this way we can generate both prediction intervals, conditional upon given parameter estimates, and prediction intervals also capturing estimation inaccuracy.

that under the identifying assumption $\mu_u = 0$, it makes sense to take $\widehat{\mu}_u = 0$ and $\widehat{a} = \widehat{A}$ (using $A = a + B\mu_u$). Thus, imposing $\mu_u = 0$, we can estimate the long run trend (a) by the estimated sample trend ($\widehat{A}$). However, we can also obtain as alternative estimator for μ_u

$$\widehat{\mu}_u = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t = \frac{1}{T}\sum_{t=1}^{T}\widehat{z}_t. \tag{3.34}$$

In finite samples this alternative estimator's deviation from zero might reflect the model's difficulty in estimating the long run trend a. Instead of estimating the long run trend by the sample trend $\widehat{a} = \widehat{A}$ (when $\mu_u = 0$), we might then alternatively estimate the long run trend by

$$\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u, \tag{3.35}$$

using the alternative estimator for μ_u. In this way we might obtain an estimate of the long run trend that is somewhat less sensitive to volatility in the mortality data than the estimator $\widehat{a} = \widehat{A}$. In the empirical analysis of the next section we shall investigate both ways of estimating the long run trend.

3.4 Empirical analysis

3.4.1 Data

We use 154 yearly observations of age-specific death numbers (D_{xt}) and exposures (E_{xt}) for men in the Netherlands, from 1850 till 2003, provided by The Human Mortality Database.[7] As in Lee and Carter (1992), we create the following age-groups: 1-4, 5-9, 10-14,...80-84, and 85+. Since the database provides data starting at the middle of the 19th century, and the number of people in age-groups above 85 (for example, 85-89, or 90-94, etc.) is relatively low in that period, we merge all the age groups above 85, resulting in the 85+ category.[8]

[7]Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 01.12.2004).

[8]Alternatively, assumptions on old-age mortality could be imposed (see for example, Coale and Guo, 1989).

3.4.2 Further specifications

In the empirical application, we impose some additional structure. For identification purposes we impose that the covariance matrix of the vector of error terms ψ_t is the identity matrix of dimension $nf \times nf$, and that the matrices Γ and Ξ appearing in the equation for the underlying latent vector u_t are restricted to be lower triangular matrices. In addition, we impose extra structure for the covariance matrix of the vector of error terms ξ_t, and for the moving average matrix Θ:

$$\Sigma_\xi = diag\left(\sigma^2_{\xi,1}, \cdots, \sigma^2_{\xi,na}\right), \tag{3.36}$$

$$\Theta = diag\left(\Theta_1, ..., \Theta_{na}\right). \tag{3.37}$$

This additional structure imposed on Σ_ξ and Θ is also intended to ease the estimation, since this structure actually implies some overidentification constraints[9].

Moreover, with na age groups and nf latent factors, the number of parameters in a and B to be estimated equals $na \times (nf + 1)$. In order to reduce this number of parameters, and to avoid localized age-induced anomalies, we use spline interpolation (see also Renshaw and Haberman, 2003a), described in Appendix 3.A.

3.4.3 Sample period sensitivity

Before presenting the estimation results when using the whole sample, we first illustrate the effect of changing the sample period. In Table 3.1 we present for some age

[9]Out of the $\frac{na\times(na+1)}{2}$ moments in $Var(m_t)$ and the $na \times na$ moments in $Cov(m_t, m_{t-1})$, $na \times nf$ moments are used to identify B, $\frac{nf\times(nf+1)}{2}$ are used to estimate the lower triangular components of Γ, and $\frac{nf\times(nf+1)}{2}$ are used to identify the lower triangular elements of Ξ. Consequently, $na \times nf + nf \times (nf+1)$ elements of Σ_ξ and Θ cannot be estimated. If $na \geq 5$ and $nf \leq 3$ for instance, after estimating B, Γ, and Ξ, there are $\frac{1}{2}\times na\times(na+1)+na\times na-na\times nf-nf\times(nf+1)$ moments left. We use additional $2\times na$ moments to estimate the diagonal elements of Σ_ξ and Θ. The rest of the moments are not used to identify additional parameters, therefore, $\frac{1}{2}\times na\times(na+1)+na\times na-na\times nf-nf\times(nf+1)-2\times na$ of the out of the diagonal elements in Σ_ξ and Θ which could be estimated are normalized to zero. If $na = 18$ and $nf = 2$ (see Tables 3.5 and 3.6), 75 out of the 495 moments given by $Var(m_t)$ and $Cov(m_t, m_{t-1})$ are used to estimate B, Γ or Ξ, and the diagonals of Σ_ξ and Θ, therefore there are 420 overidentification constraints. In order to test the effect of overidentification, we calculated the Breusch-Pagan statistic for cross-sectional independence in the residuals ξ_t for several model specifications (see Breusch and Pagan, 1980). Even though the statistical tests indicated significant cross-sectional correlation among the residuals (weaker correlation for multiple-factor models than for the 1-factor case), we decided to estimate only the diagonal elements of Σ_ξ and Θ in order to ease the convergence of the high dimensional numerical optimization problem in the subsequent estimations.

groups in the range 40 to 69 years the estimation results for two versions of the model, namely, one with a single latent factor following an AR(1)-process (1F AR) and one with a single factor following an MA(1)-process (1F MA), where the sample is either 1850-1945 or 1850-1946, where the two end years have substantially different mortality data, due to the peak in the registered number of deaths in the year 1945. We report the estimation results for $\widehat{a} = \widehat{A}$ (under the heading A) and for $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u (under the heading a). We also present the estimates of θ according to the Lee and Carter (1992)-specification following Girosi and King (2005b) (reported under the heading A).

Age group (x)	1850-1945 LC	1850-1945 1F AR	1850-1945 1F MA	1850-1946 LC	1850-1946 1F AR	1850-1946 1F MA
		A_x			A_x	
45-49	-0.0019	-0.0027	-0.0032	-0.0121	-0.0082	-0.0089
50-54	-0.0028	-0.0033	-0.0037	-0.0109	-0.0078	-0.0084
55-59	-0.0016	-0.0024	-0.0027	-0.0091	-0.0063	-0.0069
60-64	-0.0011	-0.0017	-0.0021	-0.0075	-0.0052	-0.0057
65-69	-0.0003	-0.0010	-0.0013	-0.0066	-0.0043	-0.0048
		a_x			a_x	
45-49		-0.0034	-0.0044		-0.0046	-0.0061
50-54		-0.0039	-0.0049		-0.0047	-0.0060
55-59		-0.0029	-0.0037		-0.0037	-0.0048
60-64		-0.0022	-0.0030		-0.0028	-0.0038
65-69		-0.0014	-0.0021		-0.0021	-0.0030

Table 3.1: Estimation results of A-s and a-s for various models and age groups, using as subsamples 1850-1945 and 1850-1946. The table shows the impact of the sample period (1850-1945, or alternatively 1850-1946) on the differences between the long run trend ($\widehat{a}$) and sample trend ($\widehat{A}$) estimated by various models: Lee-Carter (LC), single factor first-order autoregressive (1F AR) and single factor first-order moving average (1F MA), for various age groups: 45-49, 50-54, etc.

Changing the sample period from 1850-1945 to 1850-1946 has a dramatic impact on the estimates of A, particularly in case of the Lee and Carter (1992)-specification. However, the corresponding estimates in terms of $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ are much more stable, indicating that this way of estimating the long run trend indeed does seem to do its job: making the estimated long-run trend less sensitive to shocks in the data.

Since the estimates of a depend on the sample estimates of μ_u, in the subsequent figure we further illustrate the sensitivity of $\widehat{\mu}_u$ to the sample period employed. In the upper panels of Figure 3.2 we estimate the 1F AR and the 1F MA models for the sample

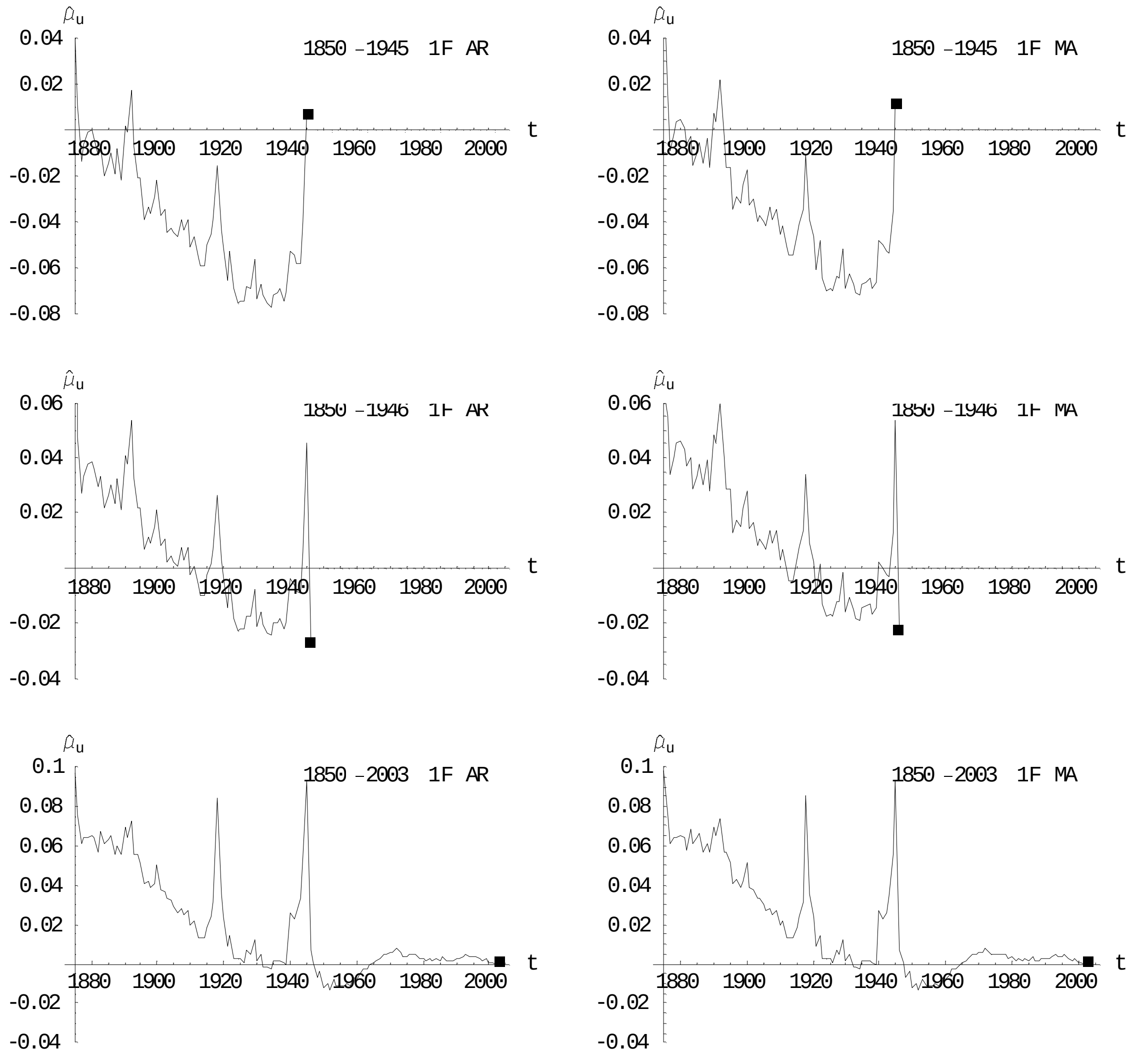

Figure 3.2: Sensitivity of $\widehat{\mu}_u$ to various sample periods. The figure shows the sensitivity of the estimated mean of the latent process ($\widehat{\mu}_u$, represented by a square at the end of the employed sample period on the figure) to different sample periods: 1850-1945, 1850-1946, and 1850-2003, with two model specifications: single factor first-order autoregressive (1F AR) and single factor first-order moving average (1F MA).

between 1850 and 1945 and report how the sample estimates of μ_u

$$\widehat{\mu}_u = \frac{1}{T-t+1}\sum_{t=1850}^{T}\widehat{u}_t \tag{3.38}$$

evolve if T runs from 1875 till 1945, based on the estimated $\widehat{u}_t$-s. This shows, that the positive shock in year 1945 has an impact: $\widehat{\mu}_u$ estimated for the period between 1850 and 1945 (denoted by the square on the figure) positively deviates from zero for both model specifications.

Similarly in the middle panels, the 1F AR and the 1F MA specifications were estimated for the sample period of 1850-1946, with a large negative shock in the registered number of deaths in year 1946. The estimates for μ_u between 1850 and 1946 for both model specifications are negative.

Finally, we considered the full sample between 1850 and 2003 in the lower panels of Figure 3.2 without any (visible) shock in year 2003. $\widehat{\mu}_u$ is close to zero both for 1F AR and 1F MA in between years 1850 and 2003, which indicates that the estimated sample trend ($\widehat{A}$) is sufficiently close to the long run trend estimate ($\widehat{a}$).

3.4.4 Estimation results

In this subsection we present the estimation results of the various model specifications using the whole available sample period.[10] We start with the Lee and Carter (1992)-benchmark, following the specification of Girosi and King (2005b). Furthermore, we estimate the model described in Section 3.3 for different specifications of the latent factor. We consider 1- and 2-factor versions, following first order autoregressive (AR) or moving average (MA) processes, and we estimate the long run trend both by means of $\widehat{a} = \widehat{A}$ and by means of $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u.

Table 3.2 contains the Lee and Carter (1992)-benchmark.[11] To ease a comparison with the other models we use the heading A to refer to the parameter vector θ. The results indicate that the decreasing trend in mortality is steepest for the youngest age group, increasing to a value close to zero (and statistically insignificant) for the oldest

[10] In Figure 3.1 describing the log mortality of men for the age group of 50-54 over time, two events that resulted in an increase of the registered number of deaths can be identified: 1) the "Spanish flu" epidemic around the year 1918, and, 2) the so-called "Dutch Hunger Winter" at the end of the Second World War. These two events affected all age groups, some of them to a larger, some of them to a smaller extent. The main reason why we did not include dummies into the time-series of the log mortality rates to filter out these events as happens sometimes in other studies is as follows. The two events increased the number of deaths particularly among the more vulnerable, with a consequence, that in the subsequent periods the mortality experience of a potentially stronger population is observed with better survival characteristics. If we had filtered out the effects of the two events, the properties of the time-series process would be potentially affected, since it reflected the mortality experience as if the population consisted of the stronger members only.

[11] There is hardly any difference between the original Lee and Carter (1992)-results without using splines and the results with splines that we employ. The restrictions imposed by the six parameter spline only reduces the log-likelihood value very marginally while the age profile is slightly smoothed.

	Coefficients		
Age group (x)	A_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.032 (0.015)	-0.007 (0.524)	0.152 (0.014)
5-9	-0.024 (0.011)	-0.030 (0.156)	0.161 (0.014)
10-14	-0.021 (0.012)	-0.162 (0.190)	0.167 (0.019)
15-19	-0.020 (0.012)	-0.084 (0.201)	0.172 (0.032)
20-24	-0.019 (0.012)	-0.335 (0.229)	0.219 (0.038)
25-29	-0.018 (0.011)	-0.353 (0.307)	0.211 (0.036)
30-34	-0.017 (0.010)	-0.387 (0.304)	0.196 (0.033)
35-39	-0.015 (0.008)	-0.360 (0.330)	0.167 (0.029)
40-44	-0.013 (0.007)	-0.400 (0.336)	0.145 (0.022)
45-49	-0.012 (0.005)	-0.396 (0.374)	0.117 (0.019)
50-54	-0.010 (0.004)	-0.490 (0.270)	0.096 (0.013)
55-59	-0.009 (0.004)	-0.501 (0.287)	0.084 (0.012)
60-64	-0.007 (0.003)	-0.452 (0.229)	0.078 (0.011)
65-69	-0.006 (0.003)	-0.506 (0.155)	0.073 (0.010)
70-74	-0.005 (0.003)	-0.511 (0.156)	0.069 (0.008)
75-79	-0.004 (0.002)	-0.598 (0.152)	0.073 (0.008)
80-84	-0.003 (0.002)	-0.645 (0.142)	0.073 (0.006)
85+	-0.002 (0.002)	-0.684 (0.176)	0.080 (0.006)
Log-Likelihood	1944.46		

Note: This table reports QML estimates and standard errors of the one-factor LC mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 3.2: 1-factor LC model with splines. The table shows the estimated parameters of the Lee-Carter model.

age group. The moving average terms are insignificant for the lower age groups, but become significantly negative for the older age groups. The standard deviations of the white noise error terms ξ_t are always substantial and estimated quite accurately, with the higher age groups having smaller standard deviations. These standard deviations are larger for the younger age groups, particularly, for the age groups 20-29 year.

In Tables 3.3 and 3.4 we present the estimation results for the 1-factor models, with Table 3.3 containing the estimation results in case the latent factor follows an AR(1) process, while in Table 3.4 the latent factor follows an MA(1) process. The heading A refers to the estimate $\widehat{a} = \widehat{A}$, while the heading a refers to the estimate $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, using the alternative estimator for μ_u. Compared to the Lee-Carter estimation results, we first notice a dramatic increase in the log-likelihood, suggesting a significant statistical improvement. The improved fit of these 1-factor models is reflected by sometimes substantially smaller estimates of the standard deviations of the error terms ξ_t compared to the Lee-Carter specifications. In addition, the structure of the error terms ξ_t also changes. Particularly, the moving average coefficients (Θ) of the lower age groups get more negative values and become significantly different from zero.

	Coefficients				
Γ	-0.244 (0.240)				
Σ_ψ	*1*				
Age group (x)	A_x	a_x	B_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.032 (0.009)	-0.032 (0.016)	0.114 (0.030)	-0.500 (0.231)	0.120 (0.009)
5-9	-0.024 (0.008)	-0.024 (0.016)	0.115 (0.020)	-0.358 (0.185)	0.116 (0.008)
10-14	-0.021 (0.010)	-0.021 (0.019)	0.143 (0.024)	-0.608 (0.121)	0.097 (0.008)
15-19	-0.020 (0.012)	-0.020 (0.023)	0.178 (0.033)	-0.467 (0.100)	0.086 (0.006)
20-24	-0.020 (0.013)	-0.019 (0.026)	0.198 (0.040)	-0.454 (0.159)	0.108 (0.017)
25-29	-0.019 (0.013)	-0.019 (0.026)	0.199 (0.042)	-0.590 (0.075)	0.077 (0.008)
30-34	-0.017 (0.012)	-0.017 (0.024)	0.185 (0.039)	-0.659 (0.056)	0.066 (0.007)
35-39	-0.016 (0.010)	-0.016 (0.021)	0.162 (0.033)	-0.723 (0.075)	0.044 (0.004)
40-44	-0.014 (0.009)	-0.014 (0.018)	0.136 (0.027)	-0.767 (0.051)	0.036 (0.002)
45-49	-0.012 (0.007)	-0.012 (0.015)	0.112 (0.022)	-0.492 (0.087)	0.035 (0.003)
50-54	-0.011 (0.006)	-0.011 (0.012)	0.093 (0.019)	-0.390 (0.097)	0.041 (0.004)
55-59	-0.009 (0.005)	-0.009 (0.010)	0.078 (0.017)	-0.362 (0.092)	0.039 (0.004)
60-64	-0.008 (0.005)	-0.008 (0.009)	0.067 (0.016)	-0.307 (0.102)	0.042 (0.004)
65-69	-0.006 (0.004)	-0.006 (0.008)	0.058 (0.015)	-0.375 (0.116)	0.044 (0.004)
70-74	-0.005 (0.004)	-0.005 (0.007)	0.052 (0.014)	-0.478 (0.112)	0.047 (0.004)
75-79	-0.004 (0.003)	-0.004 (0.006)	0.048 (0.013)	-0.700 (0.110)	0.051 (0.005)
80-84	-0.003 (0.003)	-0.003 (0.006)	0.044 (0.012)	-0.813 (0.077)	0.057 (0.005)
85+	-0.002 (0.003)	-0.002 (0.006)	0.041 (0.012)	-0.792 (0.100)	0.067 (0.005)
Log-Likelihood	3515.12				

Note: This table reports QML estimates and standard errors of the one-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 3.3: 1-factor model, AR latent factor. The table shows the estimated parameters of the model driven by a single factor autoregressive latent process.

The estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ are quite comparable to the ones according to $\widehat{A}$. So, the model seems to be able to fit the long run trend reasonably well. This is also reflected in the lower panels of Figure 3.2. However, the long run trend, as estimated by $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$, is estimated less accurately. The factor loadings (B) turn out to be very significant, revealing a hump shape, with a peak at the age groups 20-29 years. In the AR-version, the autoregression coefficient in the underlying latent process turns out to be insignificant. The same applies to the moving average term in the MA-specification.

Based on the Ljung-Box test the residuals of both the AR and the MA specifications seem to have the characteristics of a white noise. So, from this perspective there seems to be no need to include an ARMA specification.

	Coefficients				
Ξ	-0.326 (0.385)				
Σ_ψ	*1*				
Age group (x)	A_x	a_x	B_x	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.032 (0.008)	-0.032 (0.014)	0.115 (0.031)	-0.499 (0.235)	0.121 (0.010)
5-9	-0.024 (0.007)	-0.024 (0.014)	0.114 (0.020)	-0.357 (0.187)	0.116 (0.008)
10-14	-0.020 (0.008)	-0.021 (0.017)	0.142 (0.022)	-0.607 (0.123)	0.097 (0.008)
15-19	-0.020 (0.010)	-0.020 (0.020)	0.175 (0.031)	-0.466 (0.101)	0.086 (0.006)
20-24	-0.019 (0.011)	-0.020 (0.022)	0.195 (0.037)	-0.454 (0.158)	0.108 (0.017)
25-29	-0.018 (0.011)	-0.019 (0.022)	0.196 (0.038)	-0.589 (0.075)	0.077 (0.008)
30-34	-0.017 (0.010)	-0.017 (0.021)	0.182 (0.036)	-0.658 (0.056)	0.066 (0.007)
35-39	-0.015 (0.009)	-0.016 (0.018)	0.160 (0.030)	-0.722 (0.075)	0.044 (0.004)
40-44	-0.014 (0.007)	-0.014 (0.015)	0.134 (0.025)	-0.768 (0.052)	0.036 (0.002)
45-49	-0.012 (0.006)	-0.012 (0.013)	0.111 (0.020)	-0.493 (0.087)	0.035 (0.003)
50-54	-0.010 (0.005)	-0.011 (0.010)	0.092 (0.017)	-0.391 (0.097)	0.041 (0.004)
55-59	-0.009 (0.004)	-0.009 (0.009)	0.077 (0.016)	-0.362 (0.092)	0.039 (0.004)
60-64	-0.007 (0.004)	-0.008 (0.008)	0.066 (0.015)	-0.307 (0.102)	0.042 (0.004)
65-69	-0.006 (0.004)	-0.006 (0.007)	0.058 (0.014)	-0.375 (0.116)	0.044 (0.004)
70-74	-0.005 (0.003)	-0.005 (0.006)	0.051 (0.013)	-0.478 (0.112)	0.047 (0.004)
75-79	-0.004 (0.003)	-0.004 (0.005)	0.047 (0.013)	-0.700 (0.110)	0.051 (0.005)
80-84	-0.003 (0.003)	-0.003 (0.005)	0.043 (0.012)	-0.812 (0.076)	0.057 (0.005)
85+	-0.002 (0.003)	-0.002 (0.005)	0.040 (0.011)	-0.792 (0.100)	0.067 (0.005)
Log-Likelihood	3516.43				

Note: This table reports QML estimates and standard errors of the one-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 3.4: 1-factor model, MA latent factor. The table shows the estimated parameters of the model driven by a single factor moving average latent process.

In Tables 3.5 and 3.6 we present the estimation results for the 2-factor models, with Table 3.5 containing the AR(1) estimation results, and Table 3.6 the MA(1) results. In both cases the log-likelihood increases substantially compared to the corresponding 1-factor cases, but for the MA-case slightly more than for the AR-case. A likelihood ratio test reveals that the 1-factor versions are rejected against the 2-factor variants. However, the long run trend estimates remain more or less the same as in case of the 1-factor models, also in terms of their estimation accuracy, and the same applies to the moving average terms Θ. The standard deviations of the white noise error terms ξ_t decrease slightly, reflecting the better fit of the 2-factor variant. In the AR-version, the autoregression coefficient of the first latent factor turns out to be significant. The same applies to the moving average term of the first latent factor in the MA-specification. Similarly to the 1-factor case, the residuals of the 2-factor AR and MA latent processes seem to be white noise, so we did not include the ARMA specification.

	Coefficients					
Γ	-0.471 *0* (0.108) -0.021 -0.189 (0.135) (0.218)					
Σ_ψ	*1* *0* *0* *1*					
Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.031 (0.009)	-0.032 (0.018)	0.015 (0.031)	-0.124 (0.022)	-0.576 (0.117)	0.114 (0.009)
5-9	-0.024 (0.009)	-0.024 (0.017)	0.018 (0.034)	-0.120 (0.018)	-0.403 (0.136)	0.106 (0.008)
10-14	-0.020 (0.010)	-0.021 (0.021)	0.026 (0.043)	-0.146 (0.022)	-0.629 (0.067)	0.083 (0.006)
15-19	-0.020 (0.013)	-0.020 (0.026)	0.036 (0.052)	-0.178 (0.030)	-0.508 (0.070)	0.076 (0.005)
20-24	-0.019 (0.014)	-0.019 (0.028)	0.044 (0.057)	-0.196 (0.036)	-0.555 (0.252)	0.107 (0.017)
25-29	-0.018 (0.014)	-0.018 (0.028)	0.049 (0.056)	-0.193 (0.037)	-0.688 (0.078)	0.060 (0.006)
30-34	-0.017 (0.013)	-0.017 (0.026)	0.053 (0.051)	-0.176 (0.034)	-0.673 (0.061)	0.053 (0.004)
35-39	-0.015 (0.011)	-0.015 (0.022)	0.055 (0.043)	-0.150 (0.029)	-0.734 (0.071)	0.038 (0.003)
40-44	-0.014 (0.009)	-0.014 (0.018)	0.056 (0.035)	-0.120 (0.024)	-0.738 (0.045)	0.037 (0.002)
45-49	-0.012 (0.007)	-0.012 (0.015)	0.057 (0.028)	-0.094 (0.021)	-0.539 (0.074)	0.035 (0.002)
50-54	-0.010 (0.006)	-0.010 (0.012)	0.057 (0.022)	-0.073 (0.019)	-0.478 (0.094)	0.033 (0.002)
55-59	-0.009 (0.005)	-0.009 (0.011)	0.058 (0.017)	-0.057 (0.018)	-0.466 (0.073)	0.031 (0.002)
60-64	-0.007 (0.005)	-0.007 (0.009)	0.059 (0.014)	-0.045 (0.017)	-0.400 (0.088)	0.025 (0.002)
65-69	-0.006 (0.004)	-0.006 (0.008)	0.060 (0.012)	-0.036 (0.017)	-0.584 (0.102)	0.024 (0.002)
70-74	-0.005 (0.004)	-0.005 (0.008)	0.061 (0.010)	-0.030 (0.017)	-0.556 (0.126)	0.021 (0.002)
75-79	-0.004 (0.004)	-0.004 (0.008)	0.063 (0.009)	-0.024 (0.017)	-0.518 (0.074)	0.025 (0.003)
80-84	-0.003 (0.004)	-0.003 (0.008)	0.065 (0.009)	-0.020 (0.017)	-0.466 (0.060)	0.035 (0.003)
85+	-0.002 (0.004)	-0.002 (0.008)	0.066 (0.010)	-0.015 (0.017)	-0.366 (0.121)	0.048 (0.004)
Log-Likelihood	4019.59					

Note: This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis.
Normalized coefficients are written with italics.

Table 3.5: 2-factor model, AR latent factor. The table shows the estimated parameters of the model driven by a 2-factor autoregressive latent process.

The two factors can already capture most of the common properties (correlation) among separate age groups in the Netherlands. Both for the AR and the MA specifications, the first factor seems to be responsible for driving the old age mortality, taking into account significant estimates of the age groups above the age of 50. We could call this factor the *old age*-factor. The second factor seems to drive the young and middle age mortality, since it affects mostly the middle-aged groups, and slightly less the younger generation; however, it does not significantly influence the mortality rates of the old age groups. So, it is a *young and middle age*-factor.

	Coefficients					
Ξ	-0.602 *0* (0.140) -0.127 -0.261 (0.118) (0.358)					
Σ_ψ	*1* *0* *0* *1*					
Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.032 (0.009)	-0.032 (0.014)	0.040 (0.064)	-0.118 (0.031)	-0.576 (0.118)	0.114 (0.009)
5-9	-0.024 (0.008)	-0.024 (0.014)	0.042 (0.066)	-0.114 (0.023)	-0.407 (0.137)	0.106 (0.008)
10-14	-0.021 (0.010)	-0.021 (0.016)	0.055 (0.083)	-0.137 (0.026)	-0.633 (0.067)	0.082 (0.006)
15-19	-0.020 (0.012)	-0.020 (0.020)	0.071 (0.103)	-0.167 (0.034)	-0.513 (0.070)	0.076 (0.005)
20-24	-0.019 (0.013)	-0.020 (0.022)	0.082 (0.114)	-0.182 (0.041)	-0.554 (0.251)	0.107 (0.017)
25-29	-0.018 (0.013)	-0.019 (0.022)	0.087 (0.112)	-0.179 (0.043)	-0.677 (0.079)	0.061 (0.006)
30-34	-0.017 (0.012)	-0.018 (0.020)	0.087 (0.101)	-0.161 (0.042)	-0.665 (0.061)	0.053 (0.004)
35-39	-0.016 (0.010)	-0.016 (0.017)	0.083 (0.085)	-0.135 (0.039)	-0.728 (0.072)	0.038 (0.003)
40-44	-0.014 (0.008)	-0.014 (0.014)	0.078 (0.067)	-0.106 (0.036)	-0.739 (0.045)	0.037 (0.002)
45-49	-0.012 (0.007)	-0.013 (0.011)	0.073 (0.051)	-0.081 (0.034)	-0.538 (0.074)	0.035 (0.002)
50-54	-0.011 (0.005)	-0.011 (0.009)	0.069 (0.038)	-0.060 (0.033)	-0.472 (0.093)	0.033 (0.002)
55-59	-0.009 (0.004)	-0.009 (0.007)	0.067 (0.029)	-0.045 (0.032)	-0.458 (0.074)	0.031 (0.002)
60-64	-0.008 (0.004)	-0.008 (0.006)	0.065 (0.022)	-0.033 (0.032)	-0.383 (0.091)	0.025 (0.002)
65-69	-0.006 (0.003)	-0.006 (0.005)	0.064 (0.017)	-0.024 (0.032)	-0.559 (0.107)	0.025 (0.002)
70-74	-0.005 (0.003)	-0.005 (0.005)	0.064 (0.014)	-0.017 (0.032)	-0.538 (0.131)	0.021 (0.002)
75-79	-0.004 (0.003)	-0.004 (0.005)	0.065 (0.011)	-0.012 (0.032)	-0.538 (0.079)	0.025 (0.003)
80-84	-0.003 (0.003)	-0.003 (0.005)	0.065 (0.009)	-0.007 (0.032)	-0.481 (0.060)	0.034 (0.004)
85+	-0.002 (0.003)	-0.002 (0.005)	0.066 (0.008)	-0.002 (0.033)	-0.379 (0.117)	0.047 (0.004)
Log-Likelihood	4024.39					

Note: This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 3.6: 2-factor model, MA latent factor. The table shows the estimated parameters of the model driven by a 2-factor moving average latent process.

Table 3.7 compares the in-sample fit of the different models based on the cumulative sum of squared deviations of 1-period ahead in-sample forecasts. In case of the time varying drifts, we only report the results using the estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$. Among the models with the time-varying drift specification, the 2-factor MA has the best performance, except for the youngest age groups, which is expected, taking into account the corresponding log-likelihood values. In the short run the LC model with a constant drift performs as well as the 2-factor MA model in case of the older age groups, better at the young age groups, but worse in the middle age groups.

We also estimated the 3-factor versions of the AR and MA models. Even though the likelihood ratio test statistic indicates that the third factor explains a significant part of the variation, the process of this third latent factor seems to follow a random walk both for the AR and the MA versions, thus, capturing essentially irregular behavior, and the factor loadings belonging to the third factor have large parameter uncertainty. This is the reason we do not report these estimates here.

	Cumulative squared deviation 1850-2003				
Age group (x)	LC	1F AR	1F MA	2F AR	2F MA
1-4	3.52	3.62	3.64	3.51	3.62
5-9	3.95	4.18	4.20	4.17	4.24
10-14	4.27	4.43	4.43	4.40	4.44
15-19	4.55	4.96	4.94	4.97	5.02
20-24	7.33	7.40	7.38	7.33	7.26
25-29	6.84	6.87	6.82	6.81	6.74
30-34	5.87	5.95	5.89	5.93	5.83
35-39	4.25	4.29	4.25	4.23	4.19
40-44	3.21	3.11	3.09	3.06	3.04
45-49	2.11	2.14	2.13	2.12	2.09
50-54	1.40	1.47	1.46	1.45	1.40
55-59	1.09	1.14	1.13	1.13	1.09
60-64	0.94	0.98	0.97	0.97	0.94
65-69	0.82	0.87	0.86	0.85	0.82
70-74	0.74	0.77	0.77	0.77	0.74
75-79	0.81	0.85	0.84	0.84	0.81
80-84	0.81	0.85	0.85	0.83	0.81
85+	0.97	1.07	1.06	1.03	0.98

Table 3.7: Model performance. The table compares the in-sample fit of different models: Lee-Carter (LC), single factor autoregressive (1F AR), single factor moving average (1F MA), 2-factor autoregressive (2F AR) and 2-factor moving average (2F MA), based on the cumulative sum of squared deviations of 1-period ahead in-sample age-specific forecasts.

3.4.5 Prediction

In Figures 3.3-3.7 we plot 95%-prediction intervals for D_{xt}/E_{xt}, the ratio of age-specific death numbers and exposures, for the various models for the age group 65-69 year, taking as future the period 2003 to 2050. We make a distinction between a 95%-prediction interval, given the Quasi Maximum Likelihood (QML) estimates, and a 95% prediction interval also including the estimation inaccuracy of the QML estimates. In case of the time varying drifts, we only report for illustrative purposes the results using the estimates $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$.

The prediction intervals of the various models, given the QML estimates, are quite comparable. However, as soon as we also include estimation inaccuracy in the confidence intervals, the confidence intervals become substantially wider. In case of the Lee and Carter (1992)-model this increase in prediction interval is not as large as in the other models, reflecting that in the Lee and Carter (1992)-specification we use the more accurately estimated $\widehat{A}$, while in the other specifications we use the less accurately but likely more robustly estimated $\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$. Using in these models also the more accurately estimated $\widehat{A}$ would have resulted in more or less the same point forecasts, but smaller confidence intervals, comparable to the Lee and Carter (1992)-case. However, by using

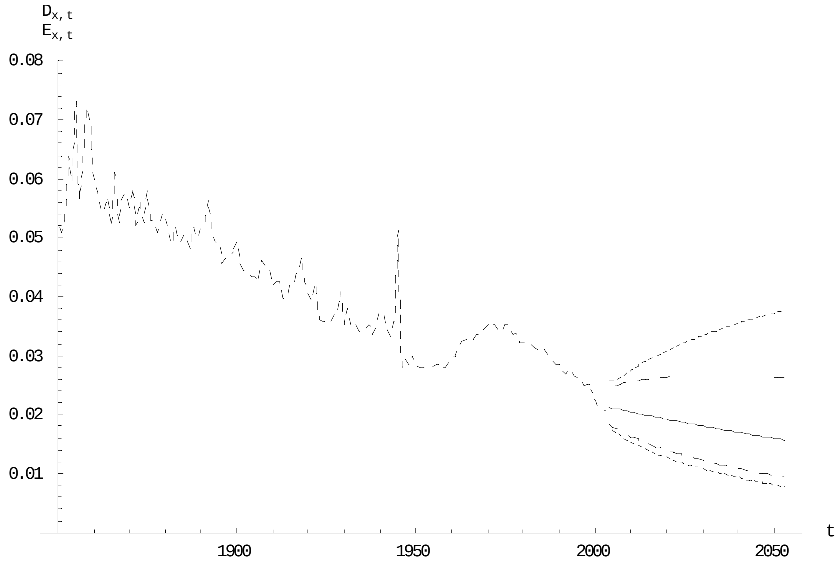

Figure 3.3: Prediction age group 65-69 year; LC spline. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the LC model for the age group of 65-69 years.

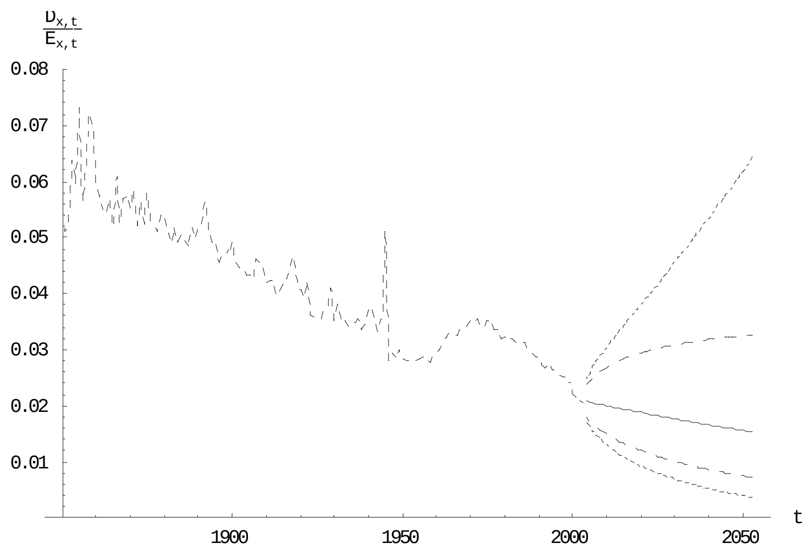

Figure 3.4: Prediction age group 65-69 year; 1-factor AR. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the 1-factor first-order autoregressive model for the age group of 65-69 years.

Figure 3.5: Prediction age group 65-69 year; 1-factor MA. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the 1-factor first-order moving average model for the age group of 65-69 years.

Figure 3.6: Prediction age group 65-69 year; 2-factor AR. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the 2-factor first-order autoregressive model for the age group of 65-69 years.

Figure 3.7: Prediction age group 65-69 year; 2-factor MA. The figure shows the prediction and the 95% prediction intervals of the ratio of age-specific death numbers and exposures, for the 2-factor first-order moving average model for the age group of 65-69 years.

$\widehat{a} = \widehat{A} - \widehat{B}\widehat{\mu}_u$ instead of $\widehat{A}$ one might incorporate the model's difficulty in estimating the long run trend. This comes at the cost of a higher inaccuracy.

3.5 Summary and conclusion

In this chapter we extended the often found empirical version of the Lee and Carter (1992) approach, as reformulated by Girosi and King (2005b). In this reformulation the log central death rates (or some other way to measure log mortalities) are directly modeled as random walks with drift. These drifts determine the long run forecasts. However, the estimation of these drifts might be rather sensitive to the sample period employed. We extended this approach by allowing for a time-varying trend, depending upon a few underlying latent factors, in order to capture the comovements between the various age groups. We formulated our model in a state-space framework, so that the Kalman filtering technique can be used to estimate the parameters by means of Quasi Maximum Likelihood.

We illustrated our specification using Dutch mortality data over the period 1850-2003. In particular we illustrated how our approach might yield a more stable estimation for the long run trend, by incorporating the model's difficulty in estimating the long run trend. When using the whole available sample period, we found comparable estimates for the trend based upon the various approaches, indicating that this sample seems to be 'representative' for the long run trend. However, since estimating the long run trend is harder when incorporating the model's difficulty in estimating the long run trend than simply estimating the long run trend by the sample trend, the prediction intervals based on the first estimated long run trends are wider than those obtained by the second approach.

3.A Appendix on spline interpolation

In this appendix we describe the spline interpolation method that we employ in order to reduce the number of parameters and to avoid localized age-induced anomalies.

i) First, let $\widetilde{x}_j$ denote the average age for age group j, $j = 1, ..., na$, where the age groups are in increasing order. Then the set $\{\widetilde{x}_1, ..., \widetilde{x}_{na}\}$ is mapped to $[0, 1]$ in the following way:

$$\{\widetilde{x}_1, ..., \widetilde{x}_{na}\} \ni \widetilde{x}_j \to \overline{x}_j = \frac{\widetilde{x}_j}{\widetilde{x}_{na}} \in [0, 1], \qquad j = 1, ..., na. \tag{3.39}$$

ii) For a we can define the function

$$F^a : \{\overline{x}_1, ..., \overline{x}_{na}\} \ni \overline{x}_j \to a_j \in \mathbb{R}. \tag{3.40}$$

Similarly, for each factor i in B, we can define the function

$$F_i^B : \{\overline{x}_1, ..., \overline{x}_m\} \ni \overline{x}_j \to B_{j,i} \in \mathbb{R}. \tag{3.41}$$

iii) The interval $[0, 1]$ is split into three intervals $[0, x^*] \cup [x^*, x^{**}] \cup [x^{**}, 1]$, where $x^* = 20/110$ divides the young-age and adult mortality, and $x^{**} = 50/110$ separates the adult and old mortality, since these have different behavior (see Heligman and Pollard, 1980), therefore, sensitivities. The functions F^a and F_i^B are approximated by cubic spline functions. For example, F^a is approximated by $\widetilde{F}^a$, where:

$$\begin{aligned} \widetilde{F}^a(x) &= S_l^a(x) \quad \text{if } x \in [0, x^*], \\ &= S_c^a(x) \quad \text{if } x \in [x^*, x^{**}], \\ &= S_r^a(x) \quad \text{if } x \in [x^{**}, 1]. \end{aligned} \tag{3.42}$$

where

$$S_l^a(x) = l_0^a + l_1^a \times x + l_2^a \times x^2 + l_3^a \times x^3, \tag{3.43}$$

$$S_c^a(x) = c_0^a + c_1^a \times x + c_2^a \times x^2 + c_3^a \times x^3, \tag{3.44}$$

$$S_r^a(x) = r_0^a + r_1^a \times x + r_2^a \times x^2 + r_3^a \times x^3. \tag{3.45}$$

We require this approximation to satisfy the smoothness conditions:

$$\begin{aligned} S_l^a(x^*) &= S_c^a(x^*), \frac{\partial S_l^a(x^*)}{\partial x} = \frac{\partial S_c^a(x^*)}{\partial x}, \ \frac{\partial^2 S_l^a(x^*)}{\partial x^2} = \frac{\partial^2 S_c^a(x^*)}{\partial x^2}, \\ S_c^a(x^*) &= S_r^a(x^*), \frac{\partial S_c^a(x^*)}{\partial x} = \frac{\partial S_r^a(x^*)}{\partial x} \text{ and } \frac{\partial^2 S_c^a(x^*)}{\partial x^2} = \frac{\partial^2 S_r^a(x^*)}{\partial x^2}. \end{aligned}$$

After solving the system of six equations, six parameters, for example, c_1^a, c_2^a, c_3^a, r_1^a, r_2^a, and r_3^a are uniquely determined by l_0^a, l_1^a, l_2^a, l_3^a, c_0^a, and r_0^a. Consequently, using splines with two knots implies the estimation of 6 parameters for a, and, similarly, 6 parameters per factor in B.

iv) Finally, the model can be written in terms of the parameters of the splines, i.e., a and B are defined by

$$a = \begin{pmatrix} \widetilde{F}^a(\overline{x}_1) \\ \vdots \\ \widetilde{F}^a(\overline{x}_{na}) \end{pmatrix}, \; B = \begin{pmatrix} \widetilde{F}_1^B(\overline{x}_1) & \cdots & \widetilde{F}_{nf}^B(\overline{x}_1) \\ \vdots & \ddots & \vdots \\ \widetilde{F}_1^B(\overline{x}_{na}) & \cdots & \widetilde{F}_{nf}^B(\overline{x}_{na}) \end{pmatrix}$$

Chapter 4

Longevity Risk in Portfolios of Pension Annuities

4.1 Introduction

The current EU solvency margin requires to hold 4% of life insurance 'mathematical' reserves as solvency capital[1]. This requirement puts some insurers at a competitive disadvantage as they have more capital locked in than the risk profile of the company would imply. This situation will change with the new solvency regulation. Risk-based solvency requirements will be introduced within Pillar 2 of Solvency II. The new supervisory principles suggested by the authorities allow more room for internal models in assessing the financial situation of insurance companies. Companies either use the capital requirements laid down by the supervisory authorities, or they are replaced by capital requirements (e.g. target capital) resulting from their own risk modeling. Target capital is risk-based and grounded in a market consistent assessment.

Since the nature of the old-age pension is very close to the life insurance business, the same phenomenon can be observed in pension fund regulation as well. Some regulators are beginning to take a more sophisticated approach to evaluate the risk profile of pension funds. The Netherlands, UK and Switzerland (and probably other countries as well) already took steps to introduce risk-based capital or funding requirements to the pension system, which is in line with the Solvency II of the EU. According to the proposal of the Dutch regulator, a capital adequacy test must be performed for three different time horizons. The minimum test ensures that the accrued benefits are covered/funded by sufficient assets in the case of immediate discontinuance. The solvency test checks the

[1]For more details, see FOPI (2004)

funding requirement in a 1-year horizon. The continuity test assesses the fund's long-term prospects.

Performing a capital adequacy test for a maturity of 1-year or for a longer horizon requires aggregate risk for the portfolio of the insurance company or pension fund to be evaluated. Several sources of risk constitute to the overall risk. Stock market risk typically influences the asset side of the portfolio, while interest rate risk influences both assets and liabilities through discounting future cash flows. The focus in this paper, however, is on the risk that results from uncertainty in the annuitant's remaining lifetime. We distinguish *micro-longevity risk*, which results from nonsystematic deviations from an individuals expected remaining lifetime, and *macro-longevity risk*, which results from the fact that survival probabilities change over time. Survival probabilities have indeed decreased for each age group in the past century accompanied by the following three phenomena: i) increasing concentration of deaths around the mode at adult ages, which is denoted as the "rectangularization" of the survival function, ii) increasing mode at adult ages of the death curve, which implies the "expansion" of the survival function, and, iii) higher level and a larger dispersion of death at young ages (for more details see e.g. Pitacco, 2004; Olivieri, 2001). The change of mortality experience clearly has a direct effect on the expected lifetime of people. Expected lifetime[2] at birth for men in the Netherlands increased from 47 years to almost 76 years in the last 100 years. For women the life expectancy at birth was 50 years at the beginning of the 20th century, while it increased above 80 years in 2000. Life expectancies for other than newly born age groups also increased. The increase in the past 100 years for men with the age of 65 was almost 4.5 years, while for the 65-year-old women it was nearly 8 years.

The goal of this chapter is twofold. First, as discussed above, solvency and continuity regulations typically require that the probability of underfunding in a given time horizon is sufficiently low. A pension fund would be underfunded if its funding ratio (the ratio of the market value of the assets to the market value of the liabilities) is below one. Therefore, we first investigate the extent to which micro- and macro-longevity risk affect the probability distribution of the future funding ratio of a portfolio of annuities.

Studies by Olivieri (2001), Coppola et al. (2000, 2003a,b), Di Lorenzo and Sibillo (2002) assume a finite number of scenarios for the evolution of future survival probabilities. They find that the micro-longevity risk for an annuity portfolio (measured by

[2]The numbers in this section reflect the estimates in the Human Mortality Database, based on period life tables. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

the variance of the payoff) becomes unimportant when the size of the portfolio becomes large. In contrast, the relative size macro-longevity risk is independent of portfolio size. Since our goal is to *quantify* the uncertainty in the future funding ratio caused by longevity risk, we use a generalized 2-factor Lee-Carter model estimated on Dutch data to forecast the probability distribution of future mortality. Specifically, the 1-year difference in the age-specific log mortality rate is modeled as the sum of two age-specific coefficients multiplied by latent time-dependent factors. The latent factors capture the common movements among mortality rates over time. In order to capture particular age-specific influences that are not properly accounted for by the model, an additional error term, which is time- and age-specific, is added. Several extensions and modifications to the methodology originally developed by Lee and Carter (1992) have been proposed, e.g. Renshaw and Haberman (2003a) and Brouhns et al. (2002, 2005).

In many empirical applications the Lee-Carter approach results in a model that describes the log central death rate by means of a linear trend, where different age groups have different trends. However, due to the volatility in mortality data, the estimation of these trends, and, thus, the forecasts based on them, are rather sensitive to the sample period employed. Chapter 3 allows for time-varying trends, depending on a few underlying factors, to make the estimates of the future trends less sensitive to the sampling period. It formulates the model in a state-space framework, and uses the Kalman filtering technique to estimate it.

We use the model estimated in Chapter 3 to simulate the distribution of future mortality rates, including micro-longevity, macro-longevity and parameter risk, and determine characteristics of the probability distribution of the funding ratio in the future. In order to quantify the effect of portfolio size, we consider portfolios of different sizes with identical annuitants. In addition, we consider portfolios of different sizes for which the age and gender composition reflects that of the Dutch population beginning of 2004. We show that, in each case, uncertainty in future survival can significantly affect the probability distribution of the funding ratio. While micro-longevity risk becomes negligible for large portfolios, macro-longevity and parameter risk remain substantial.

Our second goal is to investigate possibilities to enhance solvency in the presence of micro-and macro-longevity risk. A common method to enhance solvency of a pension fund or insurance company is to keep a fraction of its assets in a buffer. Alternatively a tailor-made contract, for instance a reinsurance contract, can be bought. A stop-loss reinsurance contract, for example, can absorb all the unfavorable future scenarios and minimize the risk of insolvency to zero. Olivieri and Pitacco (2003) calculate solvency requirements for life annuity portfolios and funded pension funds. Olivieri (2002) considers a life annuity portfolio and calculates required solvency margins for an excess of loss reinsurance contract and for a stop-loss reinsurance arrangement.[3] Their analysis is based on a mortality model with a finite number of postulated scenarios for the evolution of death rates in the future. We use the full distribution of future survival rates, estimated with a 2-factor generalized Lee-Carter model, to determine the size of the buffer required to reduce the probability of insolvency to an acceptable level, and to price a stop-loss reinsurance contract. The initial funding ratio of a large pension fund has to be as high as 107.2% to 108.4% to substantially reduce the likelihood of future insolvency in a 5-year horizon, if all sources of uncertainty in the annuitant's remaining lifetime is taken into account and financial market risk is perfectly hedged. Alternatively, the price of a stop-loss reinsurance contract for a large fund with a 5-year horizon is in the magnitude of 1.6% of the initial value of the liabilities if financial market risk is perfectly eliminated.

The chapter is organized as follows. In Section 4.2, we describe the model that is used to forecast future mortality. In Section 4.3, we determine the effect of changes in mortality rates on the expected remaining lifetime of an individual. In Section 4.4, we determine the effect on the market value of pension annuities. In Section 4.5, we assess the relative importance of micro- and macro-longevity risk for funding ratio uncertainty. In Section 4.6, we discuss several possibilities to enhance the solvency of a fund. Specifically, we determine the size of the buffer required to reduce the probability of insolvency to an acceptable level, and we determine the price of a stop-loss reinsurance contract that recovers the asset value up to the level needed to meet the solvency requirements. In order to focus on longevity risk only, the expected liabilities are considered to be cash-flow matched, so that financial market risk is eliminated. In Section 4.7, we quantify the uncertainty in the funding ratio due to longevity risk in the presence of financial market risk.

4.2 Forecasting future mortality

Due to macro-longevity risk, future survival probabilities are uncertain. In this section we present a model to estimate and forecast time-dependent survival probabilities.

[3] In case of excess of loss reinsurance, the reinsurer pays the part of the annuity that exceeds a given term, e.g. the (expected terminal) age of the annuitant specified in the contract. In case of stop-loss reinsurance, the reinsurer partially or fully recovers the required portfolio reserve at a prespecified date.

The model can be seen as a generalization of the widely used mortality forecast model introduced by Lee and Carter (1992). Let us first introduce some notation.

Notation

- $p_{x,t}$ denotes the probability at time t that a person with age x will survive at least one more year;
- $\mu_{x,t}$ denotes the *force of mortality*[4] of a person with age x at time t;
- $D_{x,t}$ denotes the observed number of deaths at time t in a cohort aged x;
- $E_{x,t}$ denotes the number of person years in a cohort aged x, the so-called exposure.

We assume that for any integer age x, and any time t, it holds that:

$$\mu_{x+u,t} = \mu_{x,t}, \qquad \text{for all } u \in [0,1), \tag{4.1}$$

Then, one can verify that

$$p_{x,t} = \exp\left(-\mu_{x,t}\right). \tag{4.2}$$

A maximum likelihood estimator for the force of mortality is given by:

$$\widehat{\mu}_{x,t} = \frac{D_{x,t}}{E_{x,t}}.^{5} \tag{4.3}$$

We use the model developed in Chapter 3 to produce forecasts of $\frac{D_{x,t}}{E_{x,t}}$. Specifically, we assume that the differentiated time series $\log \frac{D_{x,t+1}}{E_{x,t+1}} - \log \frac{D_{x,t}}{E_{x,t}}$ is driven by nf latent factors, and modeled as the sum of an age-specific constant and the the product of an age-specific coefficient vector and a vector containing the latent time-varying factors. In order to capture particular age-specific influences that are not properly accounted for by the model, an additional error term, which is time- and age-specific, is added.

In order to reduce the number of parameters to be estimated, we introduce na age groups. Let us denote $m_t = \begin{pmatrix} \ln\left(\frac{D_{1,t}}{E_{1,t}}\right) & \cdots & \ln\left(\frac{D_{na,t}}{E_{na,t}}\right) \end{pmatrix}'$ for the vector of the log force

[4]The force of mortality, at time t, of an individual with age x is defined as: $\mu_{x,t} = \lim_{\triangle t \to 0} \frac{P(0 \leqslant T_{x,t} \leqslant \triangle t)}{\triangle t}$, where $T_{x,t}$ denotes the remaining lifetime of an individual with age x at time t. For more details on estimating the force of mortality by the exposure and the death number, see Gerber (1997).

[5]For more details, see Gerber (1997).

of mortality for age groups $x \in \{1, ..., na\}$ at time t. Then the time-series evolution of m_t is modeled as

$$m_t - m_{t-1} = a + B' u_t + \xi_t + \Theta \xi_{t-1}, \tag{4.4}$$

$$u_t = \mu_u + \psi_t + \Xi \psi_{t-1}, \tag{4.5}$$

with u_t an nf-dimensional vector of underlying latent factors, driving the change in the force of mortality, where $a \in \mathbb{R}^{na}$ is the long-run trend, $B \in \mathbb{R}^{na \times nf}$ are the factor loadings, $\mu_u = E(u_t) \in \mathbb{R}^{nf}$, with $\Theta \in \mathbb{R}^{na \times na}$ a matrix with unknown parameters capturing the MA-effects in the force of mortalities, and $\Xi \in \mathbb{R}^{nf \times nf}$ a matrix with unknown parameters capturing the MA-effects in the underlying latent factors. The vectors of error terms ψ_t and ξ_t are white noise, satisfying the distributional assumption

$$\begin{pmatrix} \psi_t \\ \xi_t \end{pmatrix} | \mathcal{F}_{t-1} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\psi & 0 \\ 0 & \Sigma_\xi \end{pmatrix} \right) \tag{4.6}$$

with $\Sigma_\psi \in \mathbb{R}^{nf \times nf}$ and $\Sigma_\xi \in \mathbb{R}^{na \times na}$ the unknown covariance matrices of ψ_t and ξ_t, respectively. This model is the result of a selection procedure from a broader class of models.

We use survival data for the Netherlands (NL) for men and women separately from 1850 to 2003, provided by the Human Mortality Database.[6] We create the following 18 age-groups: 1-4, 5-9, 10-14,...80-84 and 85+. Since the database provides data starting at the middle of the 19th century, and the number of people in age-groups above 85 (e.g. 85-89, or 90-94 etc.) is relatively low in that period, we merge all the age groups above 85, resulting in the 85+ category.[7] Moreover, the maximum attainable age is assumed to be 110.

The estimation results are presented in Appendix 4.A. As it was outlined in Chapter 3, in finite samples the estimator for the mean of the latent process ($\widehat{\mu}_u$) might deviate from zero, which reflects the the model's difficulty in estimating the long run trend a. In the subsequent applications we then estimated the long run trend by

$$\widehat{a} = \widehat{A} - \widehat{B} \widehat{\mu}_u. \tag{4.7}$$

For a more detailed exposition of the model, the estimation technique, and the estimation results we refer the reader to Chapter 3.

[6] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 04.11.2005).

[7] Alternatively, assumptions on old-age mortality could be imposed (see e.g. Coale and Guo, 1989).

4.3 Uncertainty in expected remaining lifetime

Let us denote $T_{x,t}$ for the remaining lifetime at time t of an individual with age x at time t, and ${}_{\tau}p_{x,t}$ for the probability that an x-year-old at time t will survive at least another τ years, i.e.

$$ {}_{\tau}p_{x,t} = p_{x,t} \cdot p_{x+1,t+1} \cdot \cdots \cdot p_{x+\tau-1,t+\tau-1}. \tag{4.8}$$

Then, conditional on survival rates up to period t, the expected curtate remaining lifetime at time t of an individual aged x at time t equals:

$$\begin{aligned} \mathbb{E}_t\left[T_{x,t}\right] &= \sum_{\tau=1}^{110-x} \mathbb{E}_t\left[1_{(T_{x,t} \geqslant \tau)}\right] && (4.9)\\ &= \sum_{\tau=1}^{110-x} \mathbb{E}_t\left[{}_{\tau}p_{x,t}\right] && (4.10)\\ &= \sum_{\tau=1}^{110-x} \mathbb{E}_t\left[\exp\left(-\sum_{s=0}^{\tau-1} \mu_{x+s,t+s}\right)\right] && (4.11)\\ &\cong \sum_{\tau=1}^{110-x} \mathbb{E}_t\left[\exp\left(-\sum_{s=0}^{\tau-1} \frac{D_{x+s,t+s}}{E_{x+s,t+s}}\right)\right], && (4.12) \end{aligned}$$

where (4.10) follows from the law of iterated expectations, and (4.12) results from the fact the force of mortality is approximated[8] by the ratio of death numbers and exposures.

In order to illustrate the size of improvements in life expectancy, in Table 4.1 we first determine the expected remaining lifetime at the age of 25, 45 and 65 for men

[8]In order to have an idea about the effect of the approximation on the expectation, we ran a simulation experiment. We considered cohorts of independent and identical male lives with age $x \in \{'5-9', '35-39', '60-64', '80-84', '85+'\}$. For a given $\mu_{x,t}$ and given the assumption in (4.1), we simulated the realized number of deaths $D_{x,t}$ and exposures $E_{x,t}$ for a large number of times in the x-year-old male cohort, with the cohort size that was observed in the Netherlands in 2003. We calculated the relative deviation between the mean of the simulated 1-year survival probabilities based on $\frac{D_{x,t}}{E_{x,t}}$, and the survival probability based on $\mu_{x,t}$ as follows: $(E[exp(-\frac{D_{x,t}}{E_{x,t}})] - exp(-\mu_{x,t}))/exp(-\mu_{x,t})$. The experiment showed that for $\mu_{x,t}$-s that are assumed to be 0.0001, 0.001, 0.01, 0.1, 0.2 implying the corresponding 1-year survival probabilities 0.9999, 0.9990, 0.9900, 0.9048, 0.8187 that closely match the survival probabilities of the age groups of '5-9', '35-39', 60-64', '80-84' and '85+', the relative deviations are $-1 \times 10^{-6}\%$, $-5 \times 10^{-5}\%$, $-5 \times 10^{-4}\%$, $-4 \times 10^{-2}\%$ and $-1.7 \times 10^{-1}\%$ respectively. In order to illustrate the precision of the estimates for the 1-year survival probabilities, we also report the 2.5% and 97.5% quantiles $\{Q(2.5\%), Q(97.5\%)\}$ of the simulated distributions relative to $exp(-\mu_{x,t})$, calculated as follows: $(Q(.) - exp(-\mu_{x,t}))/exp(-\mu_{x,t})$. The percentage relative quantiles corresponding to the given $\mu_{x,t}$-s are $\{-0.003\%, 0.003\%\}$, $\{-0.008\%, 0.008\%\}$, $\{-0.031\%, 0.031\%\}$, $\{-0.244\%, 0.163\%\}$, and $\{-0.547\%, 0.213\%\}$ respectively.

and women, for selected historical years under the assumption that there is no further improvement in mortality, i.e. $\mu_{x+s,t+s} = \mu_{x+s,t}$, so that

$$ {}_{\tau}p_{x,t} = \exp\left(-\sum_{s=0}^{\tau-1} \mu_{x+s,t} \right) . \tag{4.13} $$

Gender	Year/Age	25	45	65
Men	1900	39.6	24.0	10.5
	1925	45.3	27.6	12.3
	1950	48.4	29.7	13.6
	1975	47.7	28.7	13.0
	2000	51.2	32.0	14.9
Women	1900	40.6	25.3	11.0
	1925	45.4	28.2	12.8
	1950	49.8	30.9	14.1
	1975	53.5	34.2	16.7
	2000	56.0	36.6	18.9

Table 4.1: Expected remaining lifetime based on period life tables. The table shows the expected remaining lifetime at the age of 25, 45 and 65 for men and women, for selected historical years under the assumption that there is no further improvement in mortality.

Next, we calculate the expected life expectancy when improvements is survival rates are taken into account. For certain age cohorts, e.g. the 25-year-old in 1900, all relevant death numbers and exposures have been observed (all the members of that cohort passed away), so that there is no randomness with respect to death rates, and life expectancy can be readily calculated. For most cohorts (that have not reached the maximum attainable age[9] in 2004), however, death numbers and exposures are needed for time periods beyond 2003, so that forecasting is required. When the forecasts of the ratio of death numbers and exposures in (4.12) are calculated, we allow for randomness in (4.4) and (4.5), e.g. ξ_t and ψ_t have nonzero variance. We note, that the resulting expected life expectancy in (4.12) for periods $t \leq 2004$ is a single number, however, the expectation of life expectancy conditional on future time periods (e.g. in 2025) is a random variable.

An additional source of risk we included in the calculations is the parameter risk, which measures the uncertainty related to the estimated parameters in (4.4) and (4.5). The presence of the parameter risk is going to yield randomness in the expected life expectancy of all the cohorts, which have not reached the maximum attainable age in year 2004.

[9]The highest attainable age is assumed to be 110.

Table 4.2 presents the expected remaining lifetime with the 95% confidence interval for parameter risk for historical time periods from 1900 to 2000. Moreover, for the expected remaining lifetime in 2025, we present two intervals. The narrower represents the 95% prediction interval of the expected life expectancy, while the wider combines the random expected life expectancy with the parameter risk.

Gender	Year/Age	25		45		65	
Men	1900	44.3		25.9		11.2	
	1925	46.4	(46.4;46.4)	28.8		12.1	
	1950	49.2	(49.1;49.3)	29.0	(29.0;29.0)	13.9	
	1975	51.9	(50.7;53.3)	31.0	(30.8;31.3)	13.4	(13.4;13.4)
	2000	53.3	(50.0;57.2)	33.1	(31.5;35.1)	15.4	(14.9;15.8)
	2025	54.6	(51.3;57.5) (39.2;64.3)	34.3	(31.6;37.0) (25.8;41.8)	16.1	(14.1;18.1) (11.1;21.2)
Women	1900	44.7		27.0		11.8	
	1925	51.1	(51.1;51.1)	29.7		12.4	
	1950	55.3	(55.1;55.7)	34.8	(34.8;34.8)	15.2	
	1975	57.2	(55.6;59.5)	36.6	(36.1;37.2)	18.2	(18.2;18.2)
	2000	58.9	(55.4;63.7)	38.2	(36.2;41.0)	19.4	(18.7;20.4)
	2025	60.6	(57.8;63.3) (52.2;70.1)	39.8	(37.2;42.4) (32.7;47.4)	20.6	(18.5;22.8) (15.8;26.0)

Table 4.2: Expected remaining lifetime based on cohort life tables. The table presents the expected remaining lifetime with the 95% confidence interval for parameter risk for historical time periods from 1900 to 2000. For the expected remaining lifetime in 2025, we present two intervals. The narrower represents the 95% prediction interval of the expected life expectancy due to the fact that the future conditional expectation is a random variable, while the wider combines the random expected life expectancy with the parameter risk.

From Table 4.1 we see that, if there had been no improvement in life expectancy after 1975, the expected remaining lifetime of a 25-year-old man would have been 47.7 years. However, if improvement is also taken into account the same cohort has a life expectancy of 51.9 years (Table 4.2). The forecasted expected remaining lifetime for a 25-year-old man in 2025 equals 54.6, with an upperbound of 57.5 without parameter risk and 64.3 with parameter risk. These results show that methods based on period life tables seriously underestimate life expectancy.

4.4 Effect of longevity on market value of annuities

We consider the market value of an annuity that guarantees a nominal yearly payment of 1, starting at the end of the year in which the annuitant reaches the age of 65, with a last payment in the year he dies. We assume that mortality risk and financial market risk are independent under the risk-neutral measure[10], and that the price of longevity risk is zero.[11] We denote $T_x = T_{x,0}$ for the current remaining lifetime of an individual with age x. Then, the current market value of an annuity for an x-year-old equals:

$$a_x = \sum_{\tau=\max\{65-x,0\}}^{110-x} \mathbb{E}\left[1_{(T_x \geqslant \tau)}\right] P_0^{(\tau+1)}, \tag{4.14}$$

where $\mathbb{E}\left[1_{(T_x \geqslant \tau)}\right]$ denotes the expected value of one unit to be paid if the annuitant is still alive at time τ, and $P_0^{(\tau)}$ denotes the market value of a zero-coupon bond maturing at time τ.

It follows from (4.12) that:

$$\mathbb{E}\left[1_{(T_x \geqslant \tau)}\right] \cong \mathbb{E}\left[\exp\left(-\sum_{s=0}^{\tau-1} \frac{D_{x+s,s}}{E_{x+s,s}}\right)\right]. \tag{4.15}$$

Now, (4.15) can be simulated by means of (4.4) and (4.5). To determine the market value of the annuity, it now only remains to specify the term structure of interest rates at $t = 0$. We will use the term structure of interest rates implied by the model presented in Subsection 4.5.2.

	Men		Women	
Age	Period Table	Projected Table	Period Table	Projected Table
25	0.872	0.944	1.038	1.139
30	1.193	1.279	1.418	1.541
35	1.633	1.733	1.939	2.086
40	2.238	2.350	2.654	2.827
45	3.079	3.198	3.643	3.840
50	4.255	4.373	5.023	5.240
55	5.918	6.022	6.950	7.177
60	8.279	8.356	9.606	9.831
65	10.403	10.441	11.969	12.179
70	8.669	8.677	10.333	10.508
75	6.897	6.881	8.490	8.617
80	5.191	5.151	6.535	6.593
85	3.723	3.675	4.643	4.680

Table 4.3: Market value of annuities. The table shows the market value of the annuity, as a function of age, based on period tables (first column), and based on forecasted mortality rates (second column), for men and for women.

Table 4.3 shows the market value of the annuity, as a function of age, for ages varying from 25 to 85 based on period tables (first column), and based on forecasted

[10] There might be some correlation between mortality and financial market factors, however we think it is negligible.

[11] This assumption is quite common in the literature. See e.g. Schrager (2006).

mortality rates (second column), for men and for women. When forecasts are made, the uncertainty in future mortality rates is taken into account (e.g. we allow for randomness in (4.4) and (4.5)). Note that the market value of the annuity increases with age until the age of 65, and starts to decrease after that. This is due to the fact that for individuals that are not yet retired, the probability that they will reach retirement increases when they get older. Moreover, discounting plays a more important role for the young. Once a person has reached retirement age, the market value of the remaining pension payments obviously decreases with age. In comparing the first and the second columns, we see that the market value of the annuity based on period life tables underestimates[12] the annuity value based on forecasted death probabilities by 7.7% for a 25-year-old man and 8.8% for a 25-year-old woman. For the 65-year-old, the corresponding numbers are 0.4% and 1.7%, respectively.

4.5 Effect of longevity risk on funding ratio uncertainty

In this section we investigate the effect of micro- and macro-longevity risk on the probability distribution of the funding ratio in the future. The funding ratio at time T (FR_T) is defined as the market value of the assets at time T (A_T) divided by the market value of the liabilities at time T (L_T), and can be seen as a measure of solvency. Formally, FR_T is defined as follows:

$$FR_T = \frac{A_T}{L_T}. \tag{4.16}$$

4.5.1 Fund characteristics

In the following analysis we consider two types of annuity funds: i) an annuity fund consisting of 65-year-old who are about to annuitize their wealth at retirement, and ii) a representative fund, which age and gender composition is the portrayal of the Dutch population at the beginning of 2004.[13] In both cases, we choose the retrospective

[12]There are some exceptions for elderly men. Due to the MA structure in (4.4) and (4.5), the predicted level of log mortality for elderly men in 2004 is higher than the level estimated for 2003, at the end of the sample period. The level correction, and the fact that discounting amortizes the effect of longer term mortality improvement, and the relatively short time horizon for mortality improvement for the elderly yield a lower annuity value with projected life table, than with period life table.

[13]CBS Netherlands, see Appendix 4.B for more details.

approach, i.e. there are no new entrants into the fund, and no rights are built up or premiums are paid after time $t = 0$. Furthermore, we assume that the maximum attainable age is $x_T = 110$, that all participants enter at $x_0 = 25$ and retire when they become 65. Consequently they contribute to the fund for maximum 40 years. We consider a nominal defined benefit fund where the right built up by a policyholder increases linearly with the amount of time he/she spent contributing to the fund, i.e. an annuitant with age x at time $t = 0$ has built up the right to receive a yearly payment of

$$D_x = \min\left\{\frac{x - x_0}{40}, 1\right\} * Q \tag{4.17}$$

after retirement, where Q denotes the yearly nominal pension payment to a person who participated for 40 years in the fund.

4.5.2 Market value of assets and liabilities

The market value of the liabilities at time T is the sum of the present value of the future cash flow stream over all individuals who are still alive at time T. Let us denote I for the initial number of participants in the fund, x_i for the age at time $t = 0$ of participant i, and T_{x_i} for the current remaining lifetime of participant i, $i = 1, \ldots, I$. We assume that mortality risk and financial market risk are independent, and that the price of longevity risk is zero. Then, the market value of the pension fund's liabilities at time T is given by:

$$L_T = \sum_{i=1}^{I} 1_{(T_{x_i} \geqslant T)} \sum_{\tau=\max\{65-(x_i+T),0\}}^{110-(x_i+T)} \mathbb{E}_T\left[{}_\tau p_{x_i+T,T}\right] P_T^{(\tau+1)} D_{x_i}, \tag{4.18}$$

where $1_{(T_{x_i} \geqslant T)}$ denotes the indicator function that is equal to one if participant i is still alive at time T, and zero otherwise, and $P_T^{(\tau)}$ denotes the market value, at time T, of a zero-coupon bond maturing at time $T + \tau$.

In order to focus exclusively on longevity risk, we assume that the expected liabilities are hedged with cash-flow matching; i.e. the initial asset portfolio consists of zero-coupon bonds paying out the initial expected value of future liabilities. Due to non-systematic and systematic deviations in mortality, the realized pension benefits typically deviate from the expectation (i.e. the payoff of the zero-coupon bond). We assume that the surplus is reinvested in, and the deficit is financed with, 1-year zero-coupon bonds. The value of the assets at time $t + 1$ therefore equals the value of the portfolio which earns the return on the 1-year bond between t and $t + 1$, minus realized pension payments at the end of period t, i.e.

$$A_{t+1} = A_t(1 + R_{t+1}) - \sum_{i:x_i+t\geqslant 65} 1_{(T_{x_i}\geqslant t)} D_{x_i}, \tag{4.19}$$

where R_{t+1} denotes the return on the bond portfolio between time t and $t+1$.

In order to be able to determine the probability distribution of the funding ratio, it now only remains to determine the market value of zero-coupon bonds. Assuming that mortality risk is not priced we postulate, following Campbell et al. (1997), that the 1-period nominal pricing kernel ($M^{\$}_{t+1}$) satisfies

$$-\log M^{\$}_{t+1} = \alpha + \delta r^{(1)}_t + \beta^{r^{(1)}} \varepsilon^{r^{(1)}}_{t+1}, \tag{4.20}$$

where $\alpha \in \mathbb{R}$, $\delta \in \mathbb{R}$, and $\beta^{r^{(1)}} \in \mathbb{R}$ are constants, where the 1-year rate $r^{(1)}_{t+1}$ follows a mean reverting process

$$r^{(1)}_{t+1} = \mu_r + \gamma\left(r^{(1)}_t - \mu_r\right) + \varepsilon^{r^{(1)}}_{t+1}, \tag{4.21}$$

where the mortality part is modeled by equations (4.4) and (4.5), and where the error terms satisfy

$$\begin{pmatrix} \varepsilon^{r^{(1)}}_{t+1} \\ \psi_{t+1} \\ \xi_{t+1} \end{pmatrix} |\mathcal{F}_t \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{r^{(1)}_t} & 0 & 0 \\ 0 & \Sigma_\psi & 0 \\ 0 & 0 & \Sigma_\xi \end{pmatrix}\right). \tag{4.22}$$

Using now that the time t price of any nominal time $t+1$ payoff X_{t+1} can be obtained via

$$P_t = \mathbb{E}_t\left[M^{\$}_{t+1} X_{t+1}\right], \tag{4.23}$$

we find that the time t zero-coupon bond price with time-to-maturity τ, $P^{(\tau)}_t$, is exponentially affine in the short rate $r^{(1)}_t$, i.e.

$$P^{(\tau)}_t = \mathbb{E}_t\left[M^{\$}_{t+1} \times \cdots \times M^{\$}_{t+\tau}\right] = \exp\left(-A_\tau - B_\tau r^{(1)}_t\right), \tag{4.24}$$

where A_τ and B_τ are constants that can easily be determined recursively from the underlying model parameters (see Campbell et al. (1997)).

We observe[14] the Dutch 1-year euro (previously guilder) interest rate swap middle rate between 1975 and 2004 on a yearly frequency, which is used to proxy the 1-year zero-

[14]The source of data for all interest rate related time-series is Datastream.

coupon yield[15]. By using the 1-year rate as the factor which drives the term structure and observing the 10-year yield with error[16] (proxied by the 10-year benchmark yield observed between 1979 and 2004), the first order autoregressive parameter for the short rate is estimated to be 0.75 with a mean of 5.4% p.a. and a standard deviation of 1.8%. The model implies a term premium of 1.2% on the 50-year bond.

Because of the long-term nature of the pension claims, the correct representation of the long end of the term structure is far more important than that of the short end. The model implied long rates at the beginning of 2004 were below the observed long rates. To fix this problem we do the following:

1. We use an equivalent representation of the term structure model using a rotation of the underlying factor, where the nominal 10-year yield ($r_t^{(10)}$) replaces the role of the 1-year interest rate ($r_t^{(1)}$). The 10-year rate follows a mean reverting process[17]

$$r_{t+1}^{(10)} = \mu_{r^{(10)}} + \gamma^{r(10)} \left(r_t^{(10)} - \mu_{r^{(10)}} \right) + \varepsilon_{t+1}^{r(10)}, \tag{4.25}$$

   with parameters $\mu_{r^{(10)}}$, $\gamma^{r^{(10)}}$, $\varepsilon_{t+1}^{r^{(10)}} \sim N\left(0, \sigma^2_{r_t^{(10)}}\right)$, which can be uniquely calculated from the parameters of the term structure model driven by $r_t^{(1)}$. Consequently, the financial market (and the term structure of interest rates) is now driven by the 10-year nominal yield ($r_t^{(10)}$):

$$P_t^{(\tau)} = \exp\left(-A_\tau^{(10)} - B_\tau^{(10)} r_t^{(10)} \right). \tag{4.26}$$

2. In order to fit the observed 10-year nominal yield perfectly, we recalculated α and δ in (4.20) such a way, that the model driven by the 10-year rate yields an identity

[15]The zero-coupon yield data are available for the period starting only from year 1997, which is very short to estimate its time-series properties. The euro/guilder interest rate swap market might contain some counterparty risk, however, the depth and the quality of the market in London is likely to make the counterparty risk limited. The comparison of the zero-yield with the swap rate in the period between 1997 and 2004 yielded a deviation of at most 0.1% point, also suggesting, that the swap rate is likely to be a good proxy.

[16]For more details, see Ang and Piazzesi (2003).

[17]Due to the unavailability of sufficient 10-year zero-coupon bond data, we were not able to estimate the dynamics of the 10-year yield directly. The dynamics of the 10-year yield is derived from the model driven by the 1-period yield. However, the characteristics of the longer term yields observed on the market are partly incorporated into the model (through A_τ and B_τ) driven by the 1-period yield, because we used a proxy for the 10-year yield (10-year benchmark yield observed with error) in order to estimate the term premium. Consequently, the model-implied dynamics of the 10-year yield also reflect the characteristics of the observed (proxied) 10-year yield to a certain extent.

for the 10-year yield in (4.26), e.g. $A_{10}^{(10)} = 0$ and $B_{10}^{(10)} = 10$. The reestimation yields $A_\tau^{(10)}$ and $B_\tau^{(10)}$.

The term structure of interest rates for January 2004 is illustrated on Figure 4.1.

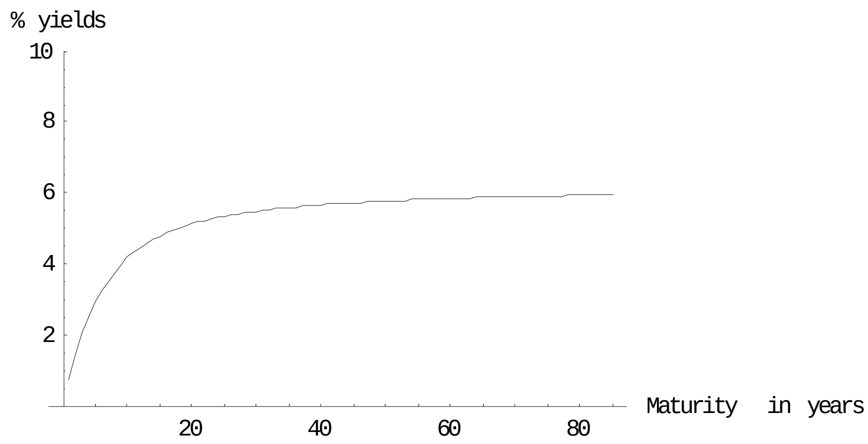

Figure 4.1: The term structure of interest rates, January 2004. The figure shows the term structure of interest rates in January 2004, if the 10-year yield is 4.2% p.a.

4.5.3 The funding ratio distribution

In this section, we quantify the effect of micro- and macro-longevity risk on the probability distribution of the funding ratio at a given time horizon. Since analytical expressions for the probability distributions of A_T and L_T are not available, we determine characteristics of the funding ratio distribution through simulation of A_T and L_T. Simulation of the value of the liabilities at time T involves:

1. Simulation of death rates for all ages and for $t = 1, \cdots, T$, using (4.4) and (4.5).

2. Simulation of $1_{(T_{x_i} \geqslant T)}$ for all participants, given the simulated death rates.

3. Determination of

$$\mathbb{E}_T \left[{}_\tau p_{x_i+T,T} \right] = \mathbb{E}_T \left[\exp \left(- \sum_{s=0}^{\tau-1} \mu_{x_i+T+s,T+s} \right) \right] \tag{4.27}$$

$$\cong \mathbb{E}_T \left[\exp \left(- \sum_{s=0}^{\tau-1} \frac{D_{x_i+T+s,T+s}}{E_{x_i+T+s,T+s}} \right) \right], \tag{4.28}$$

for every participant for which $1_{(T_{x_i} \geqslant T)} = 1$, and given the simulated death rates at time T. A closed form expression for (4.27) is not available. Because determination of (4.27) through simulation of the future value of the liabilities for every scenario generated in steps 1. and 2. is computationally intensive, we use a projection method introduced in the American option pricing literature; see e.g. Longstaff and Schwartz (2001). This method speeds up the calculations to a large extent. See Appendix 4.C for more details.

Since the asset portfolio consists of zero-coupon bonds with different maturities, the market value of the asset portfolio at time T can be simulated by means of the mortality forecast model in (4.4) and (4.5), and the term structure model in (4.24).

In the remainder of this section, we use the simulation procedure to determine characteristics of the probability distribution of the funding ratio in the future. First, in order to illustrate the effect of portfolio size, we consider portfolios of different sizes with identical annuities. Conditional on any given survival rates, the annuitants have the same survival distribution, independent of each other. In order to gain some insight into the differences between the mortality risk profiles of men and women, we consider annuity funds for 65-year-old men and annuity funds for 65-year-old women. We consider fund sizes ranging from 500 to 10,000 participants, and maturities of 1 and 5 years. In each case, the initial funding ratio is assumed to be equal to 1.

		MEN								
		Micro			Micro+Macro			Micro+Macro+Parameter		
		500	5000	10000	500	5000	10000	500	5000	10000
T=1	$StDev[FR_T]/E[FR_T]$	0.007	0.002	0.001	0.013	0.011	0.011	0.033	0.032	0.032
	Q(0.025)	0.988	0.996	0.997	0.976	0.978	0.979	0.940	0.940	0.941
	Q(0.975)	1.015	1.004	1.003	1.027	1.023	1.022	1.070	1.068	1.067
	$E[FR_T \mid FR_T < Q(0.025)]$	0.987	0.996	0.997	0.971	0.974	0.974	0.927	0.928	0.928
	$E[FR_T \mid FR_T > Q(0.975)]$	1.017	1.005	1.004	1.032	1.028	1.027	1.088	1.087	1.088
T=5	$StDev[FR_T]/E[FR_T]$	0.021	0.007	0.005	0.039	0.033	0.033	0.094	0.090	0.090
	Q(0.025)	0.962	0.988	0.991	0.929	0.940	0.940	0.841	0.848	0.851
	Q(0.975)	1.042	1.014	1.009	1.083	1.068	1.068	1.225	1.223	1.217
	$E[FR_T \mid FR_T < Q(0.025)]$	0.956	0.985	0.989	0.916	0.928	0.929	0.818	0.821	0.820
	$E[FR_T \mid FR_T > Q(0.975)]$	1.051	1.016	1.011	1.102	1.086	1.084	1.293	1.287	1.284

Table 4.4: Funding ratio distribution characteristics, 65-year-old men. The table shows the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile, the 97.5% quantile, and the expected shortfall with respect to these quantiles for an annuity portfolio, which consists of men with the age of 65 for maturities $T = 1$ and $T = 5$, for three different fund sizes (500, 5000, and 10,000), and for several (combined) risk sources (micro-, macro-longevity and parameter risk)

Table 4.4 yields the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile $Q(0.025)$, the 97.5% quantile $Q(0.975)$, and the expected shortfall with respect to these quantiles[18], for an annuity portfolio which consists of men with the age of 65 for maturities $T = 1$ and $T = 5$, and for three different fund sizes. In order to assess the relative importance of micro- and macro-longevity risk, we determine these characteristics without (columns 1-3) and with (columns 4-6) macro-longevity risk. The last three columns present the results when also parameter risk is included.[19]

In the absence of macro-longevity risk, we assume that the evolution of death rates is deterministic and given by (4.4) and (4.5) with $\xi_t = 0$ and $\psi_t = 0$, for all t. To eliminate interest rate risk, we assume that the term structure of interest rates moves deterministically to its long-term average[20], i.e. $\varepsilon_t^{R(10)} = 0$, for all t.

The riskiness in the future funding ratio increases with maturity, which is a natural consequence of the fact that the uncertainty in the time of death becomes larger. As the fund size increases, micro-longevity risk in relative terms decreases to zero, due to the pooling effect. In contrast, macro-longevity risk does not become negligible; it is almost independent of portfolio size. If parameter risk is also included in the analysis, the overall riskiness in the future funding ratio increases further.

Table 4.5 presents the results for an annuity fund of 65-year-old women. The combined micro- and macro-longevity risk is smaller for the annuity fund of 65-year-old women compared to men with the same age. If only micro- and macro-longevity risk are considered, women contribute less to the overall risk of the annuity portfolio than men. The additional risk in the parameter estimates is also larger for 65-year-old cohort of men, which is best reflected in the distribution of the future funding ratio in a 5-year horizon.

[18]Whereas pension funds are mostly interested in longevity risk, shorter than expected lifetime of the policyholders plays an important role in the risk management of life insurance companies. Our mortality model allows for improvement as well as deterioration of future survival rates. Therefore, we consider risk measures that quantify the effect of shorter than expected and longer than expected lifetime on the riskiness of the future funding ratio distribution.

[19]We draw a large number of realizations of the estimated parameters, using the robust covariance matrix of the maximum likelihood estimator. For each parameter realization, we calculate the characteristics of the funding ratio distribution, such as quantiles, variances or expected shortfall. This yields the simulated distribution of these risk measures. Depending on the risk measure, we determine either the 95% quantile or the 5% quantile of this simulated distribution.

[20]We assume that the future term structures of interest rates are in line with the implied forward rates of the today observed term structure.

		WOMEN								
		Micro			Micro+Macro			Micro+Macro+Parameter		
		500	5000	10000	500	5000	10000	500	5000	10000
T=1	$StDev[FR_T]/E[FR_T]$	0.005	0.001	0.001	0.010	0.008	0.008	0.017	0.017	0.017
	Q(0.025)	0.992	0.997	0.998	0.982	0.984	0.984	0.968	0.969	0.969
	Q(0.975)	1.010	1.003	1.002	1.020	1.017	1.017	1.035	1.033	1.034
	$E[FR_T \mid FR_T<Q(0.025)]$	0.992	0.997	0.998	0.979	0.981	0.981	0.963	0.964	0.963
	$E[FR_T \mid FR_T>Q(0.975)]$	1.011	1.004	1.003	1.024	1.020	1.020	1.040	1.039	1.038
T=5	$StDev[FR_T]/E[FR_T]$	0.014	0.004	0.003	0.029	0.025	0.025	0.039	0.036	0.036
	Q(0.025)	0.973	0.991	0.994	0.947	0.953	0.953	0.931	0.934	0.935
	Q(0.975)	1.030	1.009	1.006	1.059	1.051	1.052	1.083	1.076	1.076
	$E[FR_T \mid FR_T<Q(0.025)]$	0.969	0.990	0.993	0.938	0.945	0.946	0.919	0.923	0.924
	$E[FR_T \mid FR_T>Q(0.975)]$	1.036	1.010	1.007	1.073	1.062	1.062	1.101	1.091	1.091

Table 4.5: Funding ratio distribution characteristics, 65-year-old women. The table shows the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile, the 97.5% quantile, and the expected shortfall with respect to these quantiles for an annuity portfolio, which consists of women with the age of 65 for maturities $T = 1$ and $T = 5$, for three different fund sizes (500, 5000, and 10,000), and for several (combined) risk sources (micro-, macro-longevity and parameter risk)

Now we turn to analyze the case of a representative fund, where the age and gender composition reflects the one observed in the Dutch population at the beginning of 2004. The age and gender distribution of the Dutch population is given in Appendix 4.B. We allow for correlation between the latent processes of the mortality models which are separately estimated for men and for women. The analysis on the Dutch data implies a correlation of 0.846 between the first factors, and 0.858 between the second factors. The relatively high correlations imply that the shocks which drive the latent processes for men and women are similar. Note that the fact that the latent processes are highly correlated does not imply that the future survival probabilities of men and women move together, because i) the age-specific sensitivities are different for men and women, ii) particular age-specific error terms influence the mortality rates.

Table 4.6 presents distributional characteristics of the funding ratio for a Dutch representative fund with maturity of 1 or 5 years, and several fund sizes. The contribution of micro- and macro-longevity risk to the overall riskiness in the funding ratio is substantial. For a maturity of 5 years, micro- and macro-longevity risk imply that the standard deviation of the funding ratio is about 3.7% of its expected value for a fund with 500 participants. It decreases to 2.9% of the expected value if the fund is large (10,000 participants). Due to pooling effects, micro-longevity risk then becomes negligible. The uncertainty increases even further if parameter uncertainty of the estimates is also incorporated in the analysis. For a large fund (10,000 participants), the standard

		NL population								
		Micro			Micro+Macro			Micro+Macro+Parameter		
		500	5000	10000	500	5000	10000	500	5000	10000
T=1	$StDev[FR_T]/E[FR_T]$	0.009	0.003	0.002	0.014	0.012	0.012	0.022	0.020	0.020
	Q(0.025)	0.985	0.995	0.996	0.973	0.977	0.977	0.962	0.964	0.965
	Q(0.975)	1.019	1.006	1.004	1.029	1.024	1.024	1.044	1.040	1.040
	$E[FR_T \| FR_T<Q(0.025)]$	0.983	0.994	0.996	0.967	0.971	0.971	0.956	0.960	0.960
	$E[FR_T \| FR_T>Q(0.975)]$	1.024	1.007	1.005	1.036	1.031	1.031	1.051	1.045	1.044
T=5	$StDev[FR_T]/E[FR_T]$	0.023	0.007	0.005	0.037	0.029	0.029	0.058	0.053	0.053
	Q(0.025)	0.959	0.986	0.991	0.934	0.946	0.947	0.901	0.910	0.911
	Q(0.975)	1.048	1.015	1.010	1.077	1.060	1.058	1.120	1.113	1.113
	$E[FR_T \| FR_T<Q(0.025)]$	0.953	0.984	0.989	0.923	0.938	0.939	0.888	0.898	0.899
	$E[FR_T \| FR_T>Q(0.975)]$	1.058	1.017	1.012	1.093	1.071	1.070	1.144	1.130	1.130

Table 4.6: Funding ratio distribution characteristics, Dutch population. The table shows the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile, the 97.5% quantile, and the expected shortfall with respect to these quantiles for an annuity portfolio, which consists of an annuity population portraying the composition of the Dutch population with people older than 24. We report the risk measures for maturities $T = 1$ and $T = 5$, for three different fund sizes (500, 5000, and 10,000), and for several (combined) risk sources (micro-, macro-longevity and parameter risk)

deviation of the funding ratio in a 5-year horizon is then 5.3% of the expected value. The results show that even if uncertainty in future lifetime is the only source of risk, pension funds are already exposed to a substantial amount of uncertainty. The problem raises a hedging demand.

4.6 Management of longevity risk

Longevity bonds could potentially be used to hedge the future liabilities of a pension fund. However a longevity bond is a tool to hedge only against the macro-longevity risk, micro-longevity risk is not covered. In the previous section we saw that, as the size of the fund gets large, micro-longevity risk does not play an important role in the future uncertainty. However, for very small funds it is an important risk source. This fact already creates a mismatch between the realized liabilities of the fund and the payoff of the longevity bonds. In addition, there are other sources of mismatch related to the standardized features of longevity bonds. The longevity bonds which were issued in the UK are linked to an age group with a fixed maturity. The payoff of the product is linked to the actual evolution of a group of people (birth-cohort), which does not necessarily reflect the actual age composition of the fund. Moreover, the market for longevity bonds is very illiquid. Consequently, macro-mortality risk cannot be hedged perfectly with longevity bonds. Therefore, we analyze alternative strategies insurance companies and pension funds can use to reduce both macro- and micro-longevity risk. First, we determine the size of the buffer required to reduce the probability of underfunding to an acceptable level. Second, we determine the price of a stop-loss reinsurance contract that recovers the required portfolio reserve at a prespecified maturity. To concentrate on mortality risk we filter out all other uncertainties. Specifically, we assume that the expected liabilities are fully matched with cash-flow matching at date zero, and that the term structure of interest rates moves deterministically to its long-term average. We consider pension funds with different sizes for which the age and gender composition is that of the Dutch adult population at the beginning of 2004. In each case, the initial funding ratio is assumed to be equal to one.

4.6.1 Calibrating the solvency buffer

Solvency buffers work as an insurance mechanism because they can supplement the asset value to the level required to meet the solvency requirement at a certain maturity T.

First, we calibrate the size of the buffer such that the $VaR_{1-\varepsilon}$ (the Value-at-Risk at the $(1-\varepsilon)*100\%$ level) of the funding ratio at time T is equal to one, i.e.

$$\Pr\left(\frac{A_T+B_T}{L_T}<1\right)=\varepsilon, \tag{4.29}$$

where B_T denotes the size of the buffer at time T. We assume that the pension fund invests its buffer in a T-period risk-free zero-coupon bond, and express the buffer size at time $t=0$ as a percentage c of the initial market value of the liabilities, i.e. $B_0=cL_0$ and $B_T=cL_0/P_0^{(T)}$.

Table 4.7 presents the percentage c of the initial liability value that has to be invested in a 1- or 5-year bond (depending on the maturity) in order to meet the solvency requirement in (4.29) with $\varepsilon=0.025$.

Next, we calibrate the size of the initial buffer such that the expected shortfall of the funding ratio with respect to the $VaR_{1-\varepsilon}$ is 1 at maturity T, i.e.

$$\mathbb{E}\left[\frac{A_T+B_T}{L_T}\,\middle|\,\frac{A_T+B_T}{L_T}<VaR_{1-\varepsilon}\right]=1. \tag{4.30}$$

with $\varepsilon=0.025$.

T	N	Micro	Micro+Macro	Micro+Macro+ Parameter
T=1	500	1.455%	2.624%	3.760%
	1000	1.086%	2.412%	3.671%
	2500	0.723%	2.256%	3.582%
	5000	0.497%	2.210%	3.515%
	10000	0.358%	2.179%	3.485%
T=5	500	3.163%	5.178%	8.016%
	1000	2.331%	4.826%	7.618%
	2500	1.509%	4.486%	7.282%
	5000	1.056%	4.269%	7.281%
	10000	0.774%	4.238%	7.172%

Table 4.7: Calibrated solvency buffer, VaR. The table presents the percentage of the initial liability value that has to be invested in a 1- or 5-year bond (depending on the maturity) in order to meet the Value-at-Risk solvency requirement in (4.29) with $\varepsilon = 0.025$, with several (combined) risk sources (micro-, macro-longevity and parameter risk).

T	N	Micro	Micro+Macro	Micro+Macro+ Parameter
T=1	500	1.637%	3.190%	4.397%
	1000	1.240%	2.986%	4.241%
	2500	0.832%	2.825%	4.044%
	5000	0.596%	2.791%	3.986%
	10000	0.423%	2.784%	3.961%
T=5	500	3.792%	6.282%	9.211%
	1000	2.788%	5.774%	8.789%
	2500	1.772%	5.406%	8.492%
	5000	1.264%	5.223%	8.383%
	10000	0.909%	5.141%	8.393%

Table 4.8: Calibrated solvency buffer, expected shortfall. The table presents the percentage of the initial liability value that has to be invested in a 1- or 5-year bond (depending on the maturity) in order to meet the expected shortfall solvency requirement in (4.30) with $\varepsilon = 0.025$, with several (combined) risk sources (micro-, macro-longevity and parameter risk).

Tables 4.7 and 4.8 illustrate the importance of micro-longevity, macro-longevity and parameter risk. Depending on the risk measure (VaR or Expected Shortfall), the combination of micro- and macro-longevity risk implies that a large pension fund which is currently funded has to reserve between 4.2% and 5.1% of the initial value of the liabilities to meet the solvency requirement in a 5-year horizon. Smaller funds have to reserve even more due to the extra randomness related to micro-longevity risk. If parameter risk is included in the analysis, the initial funding ratio for large funds then has to be 107.2% and 108.4% in order to meet the solvency requirement in (4.29) and (4.30), respectively.

4.6.2 Pricing reinsurance contracts

If financial institutions offer products which hedge the risk of longevity (both macro and micro), then the market in terms of longevity risk becomes complete. A typical product that hedges both random fluctuations and macro-longevity risk is the stop-loss reinsurance contract. This reinsurance contract recovers the asset value in case the market value of the available assets is lower than the market value of the liabilities at maturity. Since the payoff of this contract is fund-specific (depends on the joint risk profile of assets and liabilities of the fund), contracts which hedge underfunding related to the uncertainty in future survival have to be tailor-made. The institution selling the reinsurance contract has to jointly model the assets and the liabilities of the fund. In this section we determine the price of the contract which takes over the longevity risk. The payoff of this contract equals:

$$Max\left(L_T - A_T, 0\right). \tag{4.31}$$

We determine the price R_0 of this contract assuming that the price of longevity risk is zero, so that:

$$R_0 = P_0^{(T)} * \mathbb{E}\left[Max\left(L_T - A_T, 0\right)\right], \tag{4.32}$$

where $P_0^{(T)}$ denotes the price of a zero-coupon bond with maturity T. As in the previous subsection, we determine the price as a percentage of the initial market value of the liabilities, i.e. $R_0 = cL_0$.

T	N	Micro	Micro+Macro	Micro+Macro+ Parameter
T=1	500	0.339%	0.539%	0.829%
	1000	0.243%	0.480%	0.805%
	2500	0.153%	0.442%	0.789%
	5000	0.106%	0.428%	0.779%
	10000	0.075%	0.423%	0.775%
T=5	500	0.689%	1.092%	1.715%
	1000	0.487%	0.987%	1.642%
	2500	0.311%	0.918%	1.609%
	5000	0.218%	0.887%	1.595%
	10000	0.155%	0.872%	1.586%

Table 4.9: Price of stop-loss reinsurance contract. The table shows the price of the reinsurance contract as a percentage of the initial value of the liabilities for maturities $T = 1$ and $T = 5$, for three different fund sizes (500, 5000, and 10,000), and for several (combined) risk sources (micro-, macro-longevity and parameter risk).

Table 4.9 shows the price of the reinsurance contract as a percentage c of the initial value of the liabilities for different fund sizes. Due to micro-, macro-longevity risk and parameter risk, the price of the reinsurance contract is at the magnitude of 1.6%-1.7% of the value of the initial liabilities at a 5-year horizon, and 0.8% at a 1-year horizon.

4.6.3 Effect on funding ratio distribution

In this section, we compare the effect of the two strategies discussed in the previous two subsections on the probability distribution of the funding ratio at maturity $T = 5$. We consider a fund with 10,000 participants. The solvency buffer required in order to decrease the probability of underfunding due to micro- and macro-longevity risk to 2.5% then equals 4.24% of the initial liability value invested in a 5-year zero-coupon bond (Table 4.7). A 5-year maturity stop-loss reinsurance contract costs 0.87% of the initial liability value (Table 4.9). To consider strategies that are comparable in terms of initial cost, we consider the following two strategies: i) a solvency buffer of 4.24% of the initial liability value invested in a 5-year zero-coupon bond, and , ii) a stop-loss reinsurance contract that supplements the asset value to the level of the liability value, if needed, combined with an additional buffer of 3.37% of the initial liability value invested in a 5-year zero-coupon bond. Then, the funding ratio at time T is given by:

$$FR_T^B = \frac{A_T}{L_T} + 0.0424\frac{L_0}{L_T P_0^{(T)}}, \text{ and,} \tag{4.33}$$

$$FR_T^R = \max\left\{1, \frac{A_T}{L_T}\right\} + 0.0337\frac{L_0}{L_T P_0^{(T)}}, \tag{4.34}$$

for the buffer and the reinsurance contract, respectively.

Figure 4.2 displays the cumulative distribution function under the two strategies. While the solvency buffer shifts up the distribution of the future funding ratio and sets the probability of underfunding to 2.5% the reinsurance contract truncates all the scenarios when the funding ratio is lower than 1, and the additional 3.37% buffer shifts up the truncated distribution.

4.7 Effect of combined longevity and market risk

In Section 4.5 we have assumed that the liabilities are matched with cash-flow matching, so that we can isolate the effect of longevity risk on funding ratio uncertainty. In this section, we include financial market risk, and determine the relative importance of micro- and macro-longevity risk in the presence of market risk.

We consider several alternative asset compositions consisting of stocks and bonds with different maturities. When dealing with stocks, we postulate that the excess stock

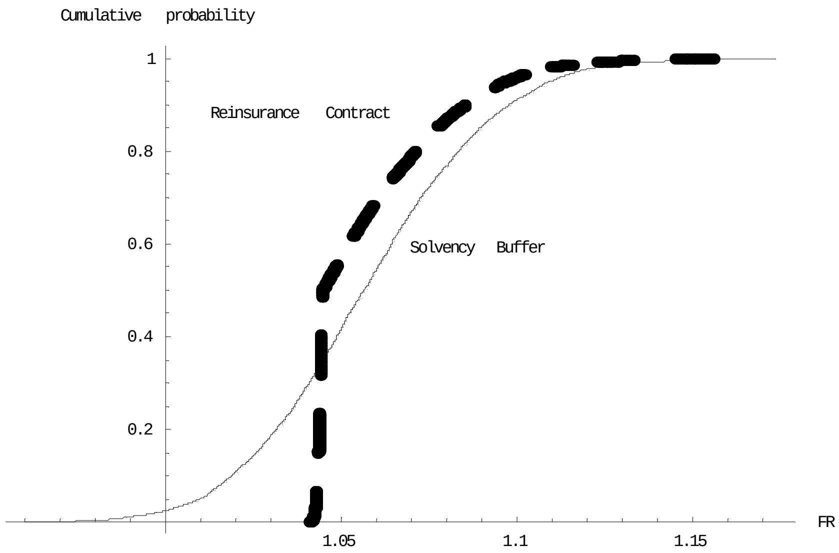

Figure 4.2: Comparison of solvency buffer and reinsurance contract. The figure shows the cumulative distribution function for a fund size of 10,000 under two equal-cost strategies. i) A solvency buffer of 4.24% of the initial liability value invested in a 5-year zero-coupon bond which sets the probability of underfunding to 2.5% in the case of micro- and macro-longevity risk, and, ii) a stop-loss reinsurance contract that supplements the asset value to the level of the liability value, if needed, combined with an additional buffer of 3.37% of the initial liability value invested in a 5-year zero-coupon bond, if micro- and macro-longevity risk are present.

return in excess of the short rate follows a random walk with drift, independently[21] of the short rate process and the mortality driving factors. This can easily be included in our market valuation model, presented in Subsection 4.5.2. We again consider pension funds of different sizes for which the age and gender composition reflects that of the Dutch population at the beginning of 2004, and assume that the initial funding ratio is one.

[21] We calculated a 0.2 sample correlation between the 1-year excess stock return and the 1-period interest rate in the period of 1985 and 2004 in the Netherlands, which is not significantly different from zero. Ang et al. (2005) documented a -0.05 correlation between the excess stock return and the 1-period short rate by using quarterly US data from 1926 and 1998, which also supports the independence assumption.

4.7.1 Data

The term structure models for the interest rates and the mortality rates we use in this section are identical to the ones estimated and introduced in the previous section. The stock market index is measured by the total return index of the Dutch market calculated by Datastream for the period between 1983 and 2004. The excess return over the short rate is estimated to be 6.2% with a volatility of 23.9% p.a.[22]

4.7.2 Uncertainty in the future funding ratio

We investigate the imperfect hedge of investment risk and its effect on the future distribution of the funding ratio combined with micro-, macro-longevity and parameter risk for five different investment strategies: i) liabilities are 'perfectly' hedged: expected liabilities are hedged with cash-flow matching initially; ii) liabilities are duration hedged, based on the McCauley duration; iii) assets are invested exclusively in 5-year bonds; iv) 50% of the assets is invested into 5-year and 50% in 10-year bonds: the interest rate elasticity of the liabilities matches the elasticity of the assets, based on the term structure model; v) 37.5% is invested into 5-year, 37.5% in 10-year bonds, and the rest is invested into stocks; vi) 25% is invested in 5-year, 25% in 10-year bonds, while the rest is invested in stocks. We investigate the 1- and 5-year horizons. The (classical) duration of the annuity portfolio is about 13 years initially, therefore the assets used to hedge the liabilities with duration matching consist of 10% 5-year and 90% 15-year bonds.

Table 4.10 and Table 4.11 show the simulated[23] distributional characteristics of the funding ratio at $T = 1$ and $T = 5$, for the above mentioned investment strategies. Because micro-longevity risk becomes negligible when the portfolio size is infinitely large, the fourth column allows to analyze the effect of different investment strategies on funding ratio uncertainty.

[22] Fama and French (2002) suggest that the equity premium estimated from fundamentals (for instance, the dividend or earnings growth rates) can be much lower than the equity premium produced by the average stock return. For simplicity, to calculate the excess return we used the average stock return in the sample from 1983 and 2004 and no fundamentals.

[23] It implies some relatively small simulation errors in some cases.

		NL population						
		Micro				Micro+Macro+Parameter		
		500	5000	10000	infinity	500	5000	10000
Perfect hedge of market risk	$StDev[FR_T]/E[FR_T]$	0.009	0.003	0.002	0.000	0.022	0.020	0.020
	Q(0.025)	0.985	0.995	0.996	1.000	0.962	0.964	0.965
	Q(0.975)	1.019	1.006	1.004	1.000	1.044	1.040	1.040
	$E[FR_T \mid FR_T<Q(0.025)]$	0.983	0.994	0.996	1.000	0.956	0.960	0.960
	$E[FR_T \mid FR_T>Q(0.975)]$	1.024	1.007	1.005	1.000	1.051	1.045	1.044
Static Duration hedge	$StDev[FR_T]/E[FR_T]$	0.014	0.012	0.011	0.011	0.024	0.023	0.023
	Q(0.025)	0.971	0.976	0.976	0.976	0.954	0.957	0.956
	Q(0.975)	1.027	1.022	1.021	1.020	1.048	1.044	1.044
	$E[FR_T \mid FR_T<Q(0.025)]$	0.965	0.971	0.971	0.972	0.949	0.952	0.953
	$E[FR_T \mid FR_T>Q(0.975)]$	1.033	1.025	1.025	1.024	1.056	1.050	1.050
100%-0%-0%	$StDev[FR_T]/E[FR_T]$	0.013	0.010	0.010	0.010	0.023	0.022	0.021
	Q(0.025)	0.978	0.982	0.983	0.983	0.959	0.962	0.963
	Q(0.975)	1.029	1.022	1.022	1.021	1.051	1.046	1.046
	$E[FR_T \mid FR_T<Q(0.025)]$	0.974	0.979	0.979	0.979	0.953	0.956	0.955
	$E[FR_T \mid FR_T>Q(0.975)]$	1.034	1.026	1.025	1.025	1.059	1.054	1.053
50%-50%-0%	$StDev[FR_T]/E[FR_T]$	0.009	0.003	0.002	0.000	0.022	0.020	0.020
	Q(0.025)	0.986	0.996	0.997	1.000	0.963	0.965	0.965
	Q(0.975)	1.020	1.007	1.005	1.001	1.045	1.042	1.041
	$E[FR_T \mid FR_T<Q(0.025)]$	0.984	0.995	0.997	1.000	0.957	0.961	0.961
	$E[FR_T \mid FR_T>Q(0.975)]$	1.025	1.008	1.006	1.001	1.052	1.046	1.045
37.5%-37.5%-25%	$StDev[FR_T]/E[FR_T]$	0.072	0.071	0.071	0.072	0.073	0.072	0.072
	Q(0.025)	0.909	0.910	0.910	0.907	0.902	0.904	0.904
	Q(0.975)	1.195	1.195	1.193	1.197	1.195	1.190	1.190
	$E[FR_T \mid FR_T<Q(0.025)]$	0.892	0.894	0.894	0.894	0.884	0.886	0.886
	$E[FR_T \mid FR_T>Q(0.975)]$	1.243	1.242	1.242	1.242	1.241	1.238	1.239
25%-25%-50%	$StDev[FR_T]/E[FR_T]$	0.139	0.138	0.138	0.141	0.137	0.136	0.136
	Q(0.025)	0.819	0.818	0.817	0.813	0.819	0.823	0.823
	Q(0.975)	1.384	1.386	1.384	1.393	1.370	1.365	1.365
	$E[FR_T \mid FR_T<Q(0.025)]$	0.786	0.787	0.787	0.787	0.789	0.789	0.789
	$E[FR_T \mid FR_T>Q(0.975)]$	1.483	1.483	1.483	1.483	1.461	1.459	1.459

Table 4.10: Distribution of future funding ratio with market risk and longevity risk combined, T=1. The table shows the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile, the 97.5% quantile, and the expected shortfall with respect to these quantiles for an annuity portfolio, which consists of an annuity population portraying the composition of the Dutch population with people older than 24. We report the risk measures for maturity $T = 1$, for several fund sizes (500, 5000, 10,000, and infinitely large fund), and for several (combined) risk sources (micro-, macro-longevity and parameter risk) under alternative investment strategies. The investment strategies are as follows: i) expected liabilities are cash-flow hedged; ii) liabilities are duration hedged ; iii) assets are invested exclusively in 5-year bonds; iv) 50% of the assets is invested into 5-year, and 50% in 10-year bonds; v) 37.5% is invested into 5-year, 37.5% in 10-year bonds, and the rest is invested into stocks; vi) 25% is invested in 5-year, 25% in 10-year bonds, while the rest is invested in stocks.

		NL population						
		Micro				Micro+Macro+Parameter		
		500	5000	10000	infinity	500	5000	10000
Perfect hedge of market risk	StDev[FR_T]/E[FR_T]	0.023	0.007	0.005	0.000	0.058	0.053	0.053
	Q(0.025)	0.959	0.986	0.991	1.000	0.901	0.910	0.911
	Q(0.975)	1.048	1.015	1.010	1.000	1.120	1.113	1.113
	E[FR_T|FR_T<Q(0.025)]	0.953	0.984	0.989	1.000	0.888	0.898	0.899
	E[FR_T|FR_T>Q(0.975)]	1.058	1.017	1.012	1.000	1.144	1.130	1.130
Static Duration hedge	StDev[FR_T]/E[FR_T]	0.038	0.032	0.031	0.031	0.069	0.065	0.064
	Q(0.025)	0.919	0.930	0.931	0.931	0.872	0.878	0.877
	Q(0.975)	1.065	1.053	1.051	1.051	1.137	1.124	1.122
	E[FR_T|FR_T<Q(0.025)]	0.903	0.916	0.916	0.917	0.854	0.863	0.864
	E[FR_T|FR_T>Q(0.975)]	1.081	1.064	1.062	1.061	1.159	1.147	1.148
100%-0%-0%	StDev[FR_T]/E[FR_T]	0.033	0.024	0.024	0.023	0.062	0.057	0.057
	Q(0.025)	0.954	0.967	0.968	0.968	0.906	0.916	0.915
	Q(0.975)	1.082	1.063	1.062	1.061	1.148	1.137	1.137
	E[FR_T|FR_T<Q(0.025)]	0.943	0.957	0.957	0.959	0.891	0.899	0.900
	E[FR_T|FR_T>Q(0.975)]	1.096	1.072	1.071	1.069	1.175	1.159	1.157
50%-50%-0%	StDev[FR_T]/E[FR_T]	0.023	0.007	0.005	0.002	0.058	0.053	0.053
	Q(0.025)	0.964	0.991	0.995	1.000	0.907	0.915	0.916
	Q(0.975)	1.054	1.021	1.016	1.009	1.129	1.119	1.120
	E[FR_T|FR_T<Q(0.025)]	0.958	0.989	0.993	0.998	0.892	0.903	0.904
	E[FR_T|FR_T>Q(0.975)]	1.064	1.024	1.019	1.010	1.152	1.138	1.137
37.5%-37.5%-25%	StDev[FR_T]/E[FR_T]	0.179	0.177	0.177	0.172	0.176	0.175	0.175
	Q(0.025)	0.825	0.825	0.825	0.832	0.819	0.826	0.826
	Q(0.975)	1.622	1.621	1.619	1.605	1.615	1.602	1.601
	E[FR_T|FR_T<Q(0.025)]	0.779	0.782	0.782	0.791	0.782	0.787	0.786
	E[FR_T|FR_T>Q(0.975)]	1.759	1.755	1.755	1.717	1.725	1.717	1.716
25%-25%-50%	StDev[FR_T]/E[FR_T]	0.346	0.346	0.345	0.335	0.333	0.331	0.331
	Q(0.025)	0.660	0.660	0.658	0.669	0.667	0.668	0.668
	Q(0.975)	2.398	2.404	2.406	2.381	2.362	2.356	2.356
	E[FR_T|FR_T<Q(0.025)]	0.586	0.587	0.587	0.601	0.604	0.608	0.608
	E[FR_T|FR_T>Q(0.975)]	2.785	2.782	2.782	2.689	2.637	2.626	2.624

Table 4.11: Distribution of future funding ratio with market risk and longevity risk combined, T=5. The table shows the standard deviation of the funding ratio relative to its expectation, the 2.5% quantile, the 97.5% quantile, and the expected shortfall with respect to these quantiles for an annuity portfolio, which consists of an annuity population portraying the composition of the Dutch population with people older than 24. We report the risk measures for maturity $T = 5$, for several fund sizes (500, 5000, 10,000, and infinitely large fund), and for several (combined) risk sources (micro-, macro-longevity and parameter risk) under alternative investment strategies. The investment strategies are as follows: i) expected liabilities are cash-flow hedged; ii) liabilities are duration hedged ; iii) assets are invested exclusively in 5-year bonds; iv) 50% of the assets is invested into 5-year, and 50% in 10-year bonds; v) 37.5% is invested into 5-year, 37.5% in 10-year bonds, and the rest is invested into stocks; vi) 25% is invested in 5-year, 25% in 10-year bonds, while the rest is invested in stocks.

If we compare the duration hedge to the asset composition of 50% 5-year and 50% 10-year bonds, we see that the relative standard deviation of the funding ratio is higher for the duration hedge. Apart from the fact that the duration of the liabilities is matched initially and the asset portfolio is not rebalanced in order to match the duration of the liabilities in the subsequent years, another reason why duration matching does not perform so well is related to the limitations of duration hedging. The interest rate sensitivity of the liabilities matches the interest rate sensitivity of the 6-year zero-coupon bond based on the term structure model we use, which explains the underperformance of the duration hedge when it is compared to the alternative bond portfolios. The 50% 5-year and 50% 10-year bond portfolio matches the interest rate sensitivity of the liabilities fairly well based on the term structure model, which explains the good hedging performance for both horizons.

The investment risk gets relatively important when the fraction of stocks increases in the asset portfolio. If we include macro-longevity and parameter risk into the analysis, then we see that the uncertainty (by looking at the relative standard deviation for instance) in the distribution of the funding ratio increases in all instances. The increase in terms of future survival uncertainty is most pronounced for the case where financial risk is perfectly hedged, and becomes relatively less important if investment risk increases. If the assets of the fund consist of 50% stocks, then the contribution of longevity risk to the overall risk of the future funding ratio becomes smaller, yet not negligible.

4.8 Conclusions

Uncertainty in the future survival probabilities contributes significantly to the riskiness of the future funding ratio. In the absence of financial market risk, a large pension fund that is currently exactly funded, and wants to reduce the probability of underfunding in a maturity of 5 years to 2.5% has to hold a buffer of about 7-8% of the initial value of the liabilities. Alternatively, longevity risk could be hedged by means of a stop-loss reinsurance contract, for which the price of a large pension fund is in the order of magnitude of 1.6%.

If market risk is also considered, the contribution of mortality risk to the overall risk of the future funding ratio becomes relatively less important. The relative importance of longevity risk decreases if the fraction of stock investments in the asset portfolio increases. However, it is not negligible.

The mortality model we considered has the potential of both future mortality improvement and mortality deterioration. We believe that we cannot exclude the risk in mortality deterioration in the future, which would significantly affect the risk of the portfolio of life insurance companies. However, given the downward sloping trend in the future, improvement is more likely than deterioration. The construction of a model which implies improvement with a large probability, and deterioration with a smaller probability is a topic for further research.

4.A Parameter estimates of the mortality model

Parameter estimates of the 2-factor moving average mortality model introduced in Section 4.2.

	Coefficients					
Ξ	-0.602 (0.140) -0.127 (0.118)	*0* -0.261 (0.358)				
Σ_ψ	*1* *0*	*0* *1*				
Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.032 (0.009)	-0.032 (0.014)	0.040 (0.064)	-0.118 (0.031)	-0.576 (0.118)	0.114 (0.009)
5-9	-0.024 (0.008)	-0.024 (0.014)	0.042 (0.066)	-0.114 (0.023)	-0.407 (0.137)	0.106 (0.008)
10-14	-0.021 (0.010)	-0.021 (0.016)	0.055 (0.083)	-0.137 (0.026)	-0.633 (0.067)	0.082 (0.006)
15-19	-0.020 (0.012)	-0.020 (0.020)	0.071 (0.103)	-0.167 (0.034)	-0.513 (0.070)	0.076 (0.005)
20-24	-0.019 (0.013)	-0.020 (0.022)	0.082 (0.114)	-0.182 (0.041)	-0.554 (0.251)	0.107 (0.017)
25-29	-0.018 (0.013)	-0.019 (0.022)	0.087 (0.112)	-0.179 (0.043)	-0.677 (0.079)	0.061 (0.006)
30-34	-0.017 (0.012)	-0.018 (0.020)	0.087 (0.101)	-0.161 (0.042)	-0.665 (0.061)	0.053 (0.004)
35-39	-0.016 (0.010)	-0.016 (0.017)	0.083 (0.085)	-0.135 (0.039)	-0.728 (0.072)	0.038 (0.003)
40-44	-0.014 (0.008)	-0.014 (0.014)	0.078 (0.067)	-0.106 (0.036)	-0.739 (0.045)	0.037 (0.002)
45-49	-0.012 (0.007)	-0.013 (0.011)	0.073 (0.051)	-0.081 (0.034)	-0.538 (0.074)	0.035 (0.002)
50-54	-0.011 (0.005)	-0.011 (0.009)	0.069 (0.038)	-0.060 (0.033)	-0.472 (0.093)	0.033 (0.002)
55-59	-0.009 (0.004)	-0.009 (0.007)	0.067 (0.029)	-0.045 (0.032)	-0.458 (0.074)	0.031 (0.002)
60-64	-0.008 (0.004)	-0.008 (0.006)	0.065 (0.022)	-0.033 (0.032)	-0.383 (0.091)	0.025 (0.002)
65-69	-0.006 (0.003)	-0.006 (0.005)	0.064 (0.017)	-0.024 (0.032)	-0.559 (0.107)	0.025 (0.002)
70-74	-0.005 (0.003)	-0.005 (0.005)	0.064 (0.014)	-0.017 (0.032)	-0.538 (0.131)	0.021 (0.002)
75-79	-0.004 (0.003)	-0.004 (0.005)	0.065 (0.011)	-0.012 (0.032)	-0.538 (0.079)	0.025 (0.003)
80-84	-0.003 (0.003)	-0.003 (0.005)	0.065 (0.009)	-0.007 (0.032)	-0.481 (0.060)	0.034 (0.004)
85+	-0.002 (0.003)	-0.002 (0.005)	0.066 (0.008)	-0.002 (0.033)	-0.379 (0.117)	0.047 (0.004)
Log-Likelihood	4024.39					

Note: This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis. Normalized coefficients are written with italics.

Table 4.12: Parameter estimates for men. The table shows the parameter estimates of the 2-factor mortality model for men in (4.4) and (4.5).

	Coefficients					
Ξ	-0.587 *0* (0.118) -0.209 -0.113 (0.113) (0.264)					
Σ_ψ	*1* *0* *0* *1*					
Age group (x)	A_x	a_x	$B_{1,x}$	$B_{2,x}$	MA(1): Θ_x	ME: $\sigma_{\xi,x}$
1-4	-0.033 (0.009)	-0.033 (0.018)	0.044 (0.02)	-0.116 (0.022)	-0.578 (0.095)	0.113 (0.009)
5-9	-0.024 (0.01)	-0.024 (0.021)	0.040 (0.021)	-0.140 (0.021)	-0.412 (0.098)	0.105 (0.006)
10-14	-0.022 (0.011)	-0.022 (0.022)	0.043 (0.025)	-0.148 (0.024)	-0.609 (0.09)	0.095 (0.007)
15-19	-0.023 (0.011)	-0.023 (0.022)	0.049 (0.028)	-0.146 (0.027)	-0.642 (0.088)	0.075 (0.007)
20-24	-0.024 (0.01)	-0.024 (0.021)	0.053 (0.029)	-0.136 (0.029)	-0.736 (0.07)	0.074 (0.006)
25-29	-0.023 (0.009)	-0.023 (0.019)	0.053 (0.027)	-0.122 (0.028)	-0.702 (0.091)	0.057 (0.005)
30-34	-0.020 (0.008)	-0.020 (0.017)	0.051 (0.022)	-0.105 (0.025)	-0.698 (0.11)	0.058 (0.007)
35-39	-0.018 (0.007)	-0.018 (0.014)	0.048 (0.017)	-0.086 (0.021)	-0.702 (0.064)	0.050 (0.005)
40-44	-0.015 (0.005)	-0.015 (0.011)	0.045 (0.013)	-0.068 (0.016)	-0.636 (0.075)	0.041 (0.003)
45-49	-0.013 (0.004)	-0.013 (0.009)	0.045 (0.01)	-0.052 (0.013)	-0.627 (0.111)	0.043 (0.004)
50-54	-0.011 (0.004)	-0.011 (0.007)	0.046 (0.008)	-0.039 (0.011)	-0.591 (0.106)	0.043 (0.004)
55-59	-0.011 (0.003)	-0.011 (0.006)	0.050 (0.007)	-0.029 (0.011)	-0.694 (0.068)	0.043 (0.004)
60-64	-0.010 (0.003)	-0.010 (0.005)	0.055 (0.006)	-0.021 (0.011)	-0.635 (0.089)	0.034 (0.004)
65-69	-0.010 (0.003)	-0.010 (0.005)	0.060 (0.007)	-0.015 (0.011)	-0.691 (0.099)	0.027 (0.002)
70-74	-0.009 (0.003)	-0.009 (0.005)	0.066 (0.007)	-0.011 (0.012)	-0.775 (0.107)	0.019 (0.002)
75-79	-0.008 (0.003)	-0.008 (0.005)	0.071 (0.008)	-0.009 (0.013)	-0.713 (0.103)	0.019 (0.003)
80-84	-0.006 (0.003)	-0.006 (0.006)	0.076 (0.009)	-0.008 (0.013)	-0.667 (0.127)	0.025 (0.004)
85+	-0.004 (0.003)	-0.004 (0.006)	0.079 (0.01)	-0.008 (0.015)	-0.519 (0.145)	0.045 (0.006)
Log-Likelihood	3961.59					

Note: This table reports QML estimates and standard errors of the two-factor affine mortality model. Standard errors are in parenthesis.
Normalized coefficients are written with italics.

Table 4.13: Parameter estimates for women. The table shows the parameter estimates of the 2-factor mortality model for women in (4.4) and (4.5).

4.B Age and gender distribution of the Dutch population

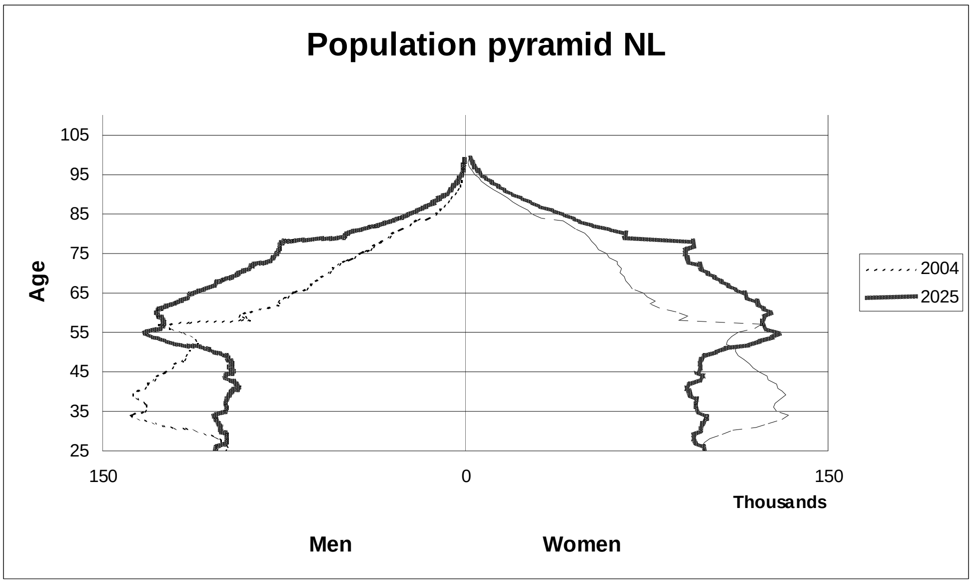

Figure 4.3: Population by age and gender in January 2004. The figure shows the population pyramid for the Netherlands at the beginning of 2004 and the expected population pyramid for 2025 for people older than 24. Men constitute 48.8% of the population with the age older than 24 in 2004. The expected number of people in the cohorts that are alive in 2025 is calculated by applying the 2-factor mortality model, which was used through the paper. Since all the cohorts of the pyramid that are older than 24 in year 2025 are alive in year 2004 already, we do not make additional assumptions on the number of newly born between 2004 and 2025. We assume there is no migration.

4.C Projection method

It is not straightforward to evaluate (4.27), since a closed form expression is not readily available. Because the determination of (4.27) through simulation for every scenario is computationally intensive, we use a projection method introduced in the American option pricing literature; see e.g. Longstaff and Schwartz (2001). This method speeds up the calculations to a large extent.

In any future scenario, $\mu_{x,t+T+l}$, $l \geq 0$, can be written as (we refer the reader to (4.4) and (4.5) and Chapter 3:

$$\mu_{x,t+T+l} = \mu_{x,t} + (T+l) * a_x + B_x \sum_{i=t+1}^{t+T+l} u_i + \sum_{i=t+1}^{t+T+l} \zeta_{x,i}. \tag{4.35}$$

If we use the information available at time $t+T$, with the remark that $\mu_{x,t+T}$ refers to the force of mortality between $t+T$ and $t+T+1$, so that it is still not known[24] at time $t+T$, then we can rewrite the previous equation as

$$\begin{aligned} \mu_{x,t+T+l} &= \mu_{x,t} + (T+l) * a_x + B_x \left(\sum_{i=t+1}^{t+T-1} u_i + \Xi \psi_{T-1} \right) \\ &\quad + \sum_{i=t+1}^{t+T-1} \zeta_{x,i} + \Theta_x \xi_{x,T-1} + \eta_{t+T+l}, \end{aligned} \tag{4.36}$$

where η_{t+T+l} captures all the unknown terms between $t+T$ and $t+T+l$:

$$\eta_{t+T+l} = B_x \left(\sum_{i=t+T}^{t+T+l} u_i - \Xi \psi_{T-1} \right) + \sum_{i=t+T}^{t+T+l} \zeta_{x,i} - \Theta_x \xi_{x,T-1}, \tag{4.37}$$

with expectation zero ($E[\eta_{t+T+l}] = 0$), which follows from the assumptions of the model.

Adding (4.36) over τ, we get

[24] It will realize only at the end of the period, at $t+T+1$, or in other words it is not locally riskless at time $t+T$ as the 1-period interest rates.

$$\begin{aligned}\sum_{i=0}^{\tau-1}\mu_{x+i,t+T+i} &= \sum_{i=0}^{\tau-1}\mu_{x+i,t}+\sum_{i=0}^{\tau-1}(T+i)*a_{x+i}+\sum_{i=0}^{\tau-1}B_{x+i}*\left(\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right)\\ &\quad+\sum_{j=0}^{\tau-1}\left[\sum_{i=t+1}^{t+T-1}\zeta_{x+j,i}+\Theta_{x+j}\xi_{x+j,T-1}\right]+\sum_{i=0}^{\tau-1}\eta_{t+T+i} \qquad (4.38)\\ &= \alpha+\beta\left(\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right)\\ &\quad+\sum_{j=0}^{\tau-1}\left[\sum_{i=t+1}^{t+T-1}\zeta_{x+j,i}+\Theta_{x+j}\xi_{x+j,T-1}\right]+\sum_{i=0}^{\tau-1}\eta_{t+T+i}, \qquad (4.39)\end{aligned}$$

where α and β are constant coefficients.

Since our main goal is to approximate ${}_{\tau}p_{x,t+T}$:

$${}_{\tau}p_{x,t+T}=\exp\left(-\sum_{i=0}^{\tau-1}\mu_{x+i,t+T+i}\right), \qquad (4.40)$$

with the variables entering into (4.38), e.g.

$$\begin{aligned}{}_{\tau}p_{x,t+T} &\simeq f\left(\begin{bmatrix}1\\0\end{bmatrix}'*\left[\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right],\begin{bmatrix}0\\1\end{bmatrix}'*\left[\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right],\right.\\ &\quad\left.\sum_{i=t+1}^{t+T-1}\zeta_{x_1,i}+\Theta_{x_1}\xi_{x_1,T-1},...,\sum_{i=t+1}^{t+T-1}\zeta_{x_n,i}+\Theta_{x_{na}}\xi_{x_n,T-1}\right)+\gamma_{t+T} \qquad (4.41)\\ &\equiv f\left(X_1,X_2,Z_1,...,Z_n\right)+\gamma_{t+T}, \qquad (4.42)\end{aligned}$$

where $X_1=\begin{bmatrix}1\\0\end{bmatrix}'*\left[\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right]$, $X_2=\begin{bmatrix}0\\1\end{bmatrix}'*\left[\sum_{i=t+1}^{t+T-1}u_i+\Xi\psi_{T-1}\right]$, $Z_1=\sum_{i=t+1}^{t+T-1}\zeta_{x_1,i}+\Theta_{x_1}\xi_{x_1,T-1}$, ... and $Z_n=\sum_{i=t+1}^{t+T-1}\zeta_{x_n,i}+\Theta_{x_{na}}\xi_{x_n,T-1}$ are basis functions, $E\left[\gamma_{t+T}\right]=0$ and $\gamma_{t+T}\in\mathbb{R}$.

Note, that we included the errors of all age groups into (4.41) which do not appear in (4.38) to simplify the implementation of the problem when constructing cohort life tables for an arbitrary age group.

Finally, we explicitly define relationship $f(.)$ on the right hand side of (4.41):

$${}_{\tau}p_{x,t+T}\simeq{}^{25}\theta\Psi+\gamma_{t+T} \qquad (4.43)$$

[25] It is interesting to look at how the specified functional form of the basis functions Ψ approximates

with $f(.) = \theta\Psi$ and $\theta \in \mathbb{R}^{1\times(8+3*18)}$ is a vector of coefficients and
$\Psi = \begin{bmatrix} 1 & X_1 & X_2 & X_1X_2 & X_1^2 & X_2^2 & X_1^3 & X_2^3 & Z_1 & \cdots & Z_n & Z_1^2 & \cdots & Z_n^2 & Z_1^3 & \cdots & Z_n^3 \end{bmatrix}'$
is the vector of scenario-specific state variables. We did not consider higher than third order terms in Ψ in order to limit the number of regressors. Scenario-specific $E_{t+T}\left[{}_{\tau}p_{x,t+T}\right]$ is calculated with projection by conditioning on the scenario-specific state variables, which yields (4.27). The projection method is extremely time efficient as opposed to the simulation method.

${}_{\tau}p_{x,t+T}$. We do not observe the force of mortalities. For the sake of simplicity we assume that $\mu_{x,t} = \frac{D_{x,t}}{E_{x,t}}$, and given the assumption in (4.1) we ran a simulation experiment. We calculated the value of a life annuity for a man with the age of 65 at time $T = 5$, i) based on the projection method: we generated $n = 2000$ scenarios to calculate the conditional expectations at $T = 5$ for each scenario and, ii) based on the simulation of the future conditional expectations at time $T = 5$: on each node of the same $n = 2000$ scenarios we ran simulations by using $k = 2000$ scenarios for every node. We calculated scenario-specific absolute value of deviations between the annuity values produced by i) and ii), and finally, we calculated the average of the absolute value of deviations over all $n = 2000$ scenarios. The average value of the absolute value of deviations is about 0.2% of the unconditional expectation of the annuity value at time $T = 5$. The estimates for the deviation are likely to contain simulation error.

Chapter 5

The Determinants of the Money's Worth of Participation in Collective Pension Schemes

5.1 Introduction

In many countries employees have implicit or explicit options to opt out of collective pension schemes. The option can be to participate in a collective pension scheme or to receive a lump-sum contribution via an individual defined contribution scheme, but it is often also more implicit. Employment with a specific firm might imply mandatory participation in the collective pension scheme of this firm, which can be avoided by switching jobs to another firm, sector or country. In this chapter we analyze the economic value or money's worth of the annuity contracts that are typically offered by collective pension schemes. Collective pension schemes are often funded via a uniform contribution, determined as a fraction of the wage earned. Therefore, the premium paid is invariant to the individual characteristics of the employee, like age, gender, and education level. For instance, the economic value of identical annuity contracts is substantially lower for young employees than for employees close to retirement due to the time value of money and a lower likelihood of surviving up to retirement. The money's worth of participation in uniformly priced pension schemes depends on other individual characteristics than age, that determine the survival probabilities of employees. It is well-known that women live longer than men in expectation, and the life expectancy of highly educated groups substantially exceeds that of lower education groups, see e.g. Brown (2003) and Huisman et al. (2004, 2005). This discrepancy between the money's worth and what employees pay introduces incentives that are analyzed in this chapter.

The collective schemes considered in this chapter can be characterized by obligatory participation, forced annuitization, collective asset allocation decisions and uniform pricing. The schemes can be either defined benefit (DB) or collective defined contribution (DC). The rights in the defined benefit plans purely depend on the labor history of the participant. In collective DC plans the asset returns and future premium rules play an important additional role, and e.g. determine whether or not the rights will be indexed against inflation. Both of the collective plans (either DB or collective DC) generate the same incentives of the participation in the schemes. Occupational earnings related collective pension plans with flat contribution rates as studied in this chapter are common in the Netherlands, UK, US, Switzerland and Canada, just to name a few.

Apart from individual heterogeneity, differences in the extent to which inflation protection is provided can also have a very significant impact on the money's worth of participation in a pension scheme. Obviously real annuities are more valuable than nominal ones, but the relative valuation in money terms will depend on the current inflation and interest rates. More importantly, many collective schemes do not offer straight nominal or real pension rights but target to provide inflation indexation if sufficient funding is available. The extent to which such schemes offer indexation protection will be referred to as the indexation quality of the scheme. In our numerical calculations we will focus on the specific indexation rules that have recently been adopted by many Dutch pension funds. In the schemes offered by these funds indexation of pension rights will typically be only partial if the funding ratio of the fund is insufficient[1]. Moreover, insufficient funding implies that subsequently the pension premium will be increased. As a consequence, employers covered by a pension fund with a currently high funding ratio are offered a more attractive pension scheme than another fund where the rules that determine the entitlements are identical but the funding is worse. Therefore, apart from the impact of individual characteristics, there is heterogeneity at the level of pension funds, which introduces additional incentives. Furthermore, we address the interplay between the individual incentives and the incentives provided by the current financial situation of the pension fund.

Analysis of the welfare effects of a pension scheme is required whenever pension schemes are evaluated or redesigned. Deviations between the cost and the market value of participation in the scheme require solidarity of groups of individuals that is not

[1] A similar indexation policy is used in Switzerland.

Pareto improving[2]. If the deviation between costs and ex ante benefits of participation would get too large, the net contributors in a voluntary scheme will not participate and the scheme may become unsustainable. Many studies have analyzed the cost and benefits of life-time participation in specific DB schemes (see e.g. Cui et al., 2005; Gollier, 2005) and considers sustainability of the scheme relative to DC schemes. This chapter in contrast focuses on costs and benefits of participation in a collective scheme for a single year.

Even if participation in a DB pension scheme is legally obliged, the net contributors will try to avoid participation and e.g. switch jobs to another firm or industry or to another country for that reason, which also makes the scheme unsustainable. Of course, the cost of switching jobs (including loss of job or sector-specific human capital) can be substantial, and the incentive to leave the fund becomes relevant only if the differential between contribution and economic value exceeds the switching costs.

Participation for one year in a DB scheme generates an annuity payment as of the retirement age. This chapter focuses on the economic value of this annuity.[3] An extensive literature has analyzed the welfare effects of annuities (see for example Brown, 2002, 2003), of holding indexed bonds (see among others Campbell and Viceira, 2001, 2002; Campbell et al., 2003; Brennan and Xia, 2002) and of other investment strategies. These papers assume specific initial assets and decision rules for investors and make a utility comparison. In line with the literature on the money's worth of annuities we restrict ourselves to the case of a fully rational optimizing agent and complete markets. We assume that the agent can and will costlessly unwind the portfolio strategy of the fund as well as of the annuities imposed by the scheme. In this setting the investment strategy of the fund, the precise form of the utility function of the agent or any additional asset holdings that the agent might have, does not have any relevance for the value of participation in the scheme. A comparison of the money's worth of the participation in the scheme and the cost of participation captures all incentives to participate.

It is well known in the literature that for subgroups of the population the money's worth of the annuities that are imposed can differ from the costs charged by uniform contributions. These differences are often referred to cost solidarity between subgroups. For Dutch schemes for instance, Kune (2005) has listed cost solidarities between men

[2]This should be distinguished from the risk solidarity that is often imposed by the premium and indexation rules in schemes which are welfare improving if equivalent financial instruments are not available.

[3]The product offered by many compulsory pension schemes also contains disability insurance and partner pension. These are not considered in the analysis.

and women, between younger and older workers, between singles and couples, between workers and disabled, between low and high educated groups, etc. Some of these solidarities might be intentional and desirable (e.g. the solidarity between workers and disabled persons), others might be non-intentional and undesirable. The aim of this chapter is to quantify the solidarities imposed by uniform pricing, not to consider which solidarities would or would not be desirable.

In our analysis we will assume that the costs of switching from one pension scheme to another are small. In reality, switching jobs can be costly and switching between funds can have a significant impact on the accumulated retirement wealth[4]. For instance, in the US and in the UK pension rights that are not yet vested can evaporate. In final wage schemes it can be rather unattractive to leave pension rights with one pension fund and switch to another since the final wage in the first scheme will no longer be adjusted. Insufficient transferability of pension rights to another scheme can imply that transfer of the funds to another scheme is likewise not too attractive. In this chapter we assume in contrast that the accumulated retirement wealth is not affected by a change of pension schemes. This is at least approximately the case in the Netherlands, where transferability of retirement funds at actuarially fair prices is a legal right. Note however, that the formulas that are currently used in Dutch pension transfers are approximations to the market values that we analyze. Moving between funds with a different indexation quality e.g. generally does have an impact on the value of retirement wealth, because the indexation quality is not taken into account. Note also that for young workers the switching costs are likely to be small in all cases, so that they have the strongest incentives to find optimal pension schemes.

Our comparison of the cost and benefits of participation in a pension scheme is closely related to generational accounting (see Auerbach et al., 1999). Ponds (2003) emphasizes the need to have ex ante fair pricing of the pension contract for each cohort. We extend his analysis to a comparison of the costs and benefits of other subgroups (men versus women and differences in education level) and to differences in indexation quality. Moreover, we analyze the incentives for every cohort year by year rather that we sum them up to an overall number.

Our emphasis in this study is on the determinants of the money's worth of participation in a collective pension scheme. We focus on the implications of the money's worth for possible options to opt out. The money's worth of participation in a pension

[4]See Cocco and Lopes (2004) for a detailed description of the rules to transfer retirement wealth between different funds in the UK.

scheme is also important for a variety of other reasons. First of all, the fair price for new entrants to a scheme (e.g. because they were recently hired by the firm that sponsors the scheme) equals the market value taking their characteristics into account. The analysis also clarifies the incentives for insurers to offer annuity products in specific segments of the population. While very few insurers do so explicitly, this can be done implicitly by focusing the marketing efforts on specific subgroups. The value of participation in a scheme is also required to have transparent labor markets where agents react to the incentives implied by the scheme. The premium to be paid for nontransparent obligatory pension schemes can easily be perceived as taxation rather than a contribution to personal income during retirement, which would imply that the net wage that is offered is underestimated and the labor market is distorted. Finally, the market value of the liabilities to all participants is also an important element in the accounting of the firm as of the introduction of the International Financial Reporting Standards. Market valuation requires that for valuation of the liabilities, the heterogeneity in the population of the pension fund is properly accounted for.

The main results of this chapter are the following. Participants in a scheme with primarily older and highly educated workers have strong incentives to opt out of a uniformly priced collective pension scheme if they have access to annuity markets at risk-based pricing. Assuming the economic conditions of January 2004 the money's worth of one year participation in a nominal scheme for a 25-year-old man is estimated to be 1.5% of his annual salary on average, while the money's worth for a 64-year-old man is 18.7%. The money's worth moreover depends significantly on gender as well as on level of education. For conditionally indexed schemes the money's worth moreover depends on he current funding ratio and on the asset mix of the fund. The money's worth of participation of a 25-year-old man in a fund with funding ratio of 100% and 100% bond investment is 1.7% of his salary, while the value of participation in an identical fund with with a funding ratio of 140% is 3.3%. We moreover show how the money's worth of participation in the scheme depends on the assumptions on improvements in life expectancies.

The set-up of this chapter is as follows. In Section 5.2 we present a review of the extensive literature on differences among groups in the welfare effects and pricing of annuities. In Section 5.3 we determine the money's worth for different groups of individuals of a year of participation in a nominal or fully indexed DB plan. The results indicate that the economic value of the annuity rights that are obtained can be substantially different, while the cost of participation is typically the same for all. We discuss the drawbacks for the cost solidarity between groups that is imposed by an obligatory collective scheme that is based on uniform pricing. Throughout, we ignore the effect of premium adjustments if the funding ratio of a fund would drop and assume that the individual could avoid this increase by switching to another employer or to third pillar products. In Section 5.4 we focus on the use of models of the nominal and real term structure similar to the ones proposed by Brennan and Xia (2002) and Campbell and Viceira (2001) to determine the market value of a year of participation in conditionally indexed schemes. Section 5.5 restates the main conclusions of the analysis.

5.2 A survey of the literature on money's worth of annuities

An extensive recent literature outlines elements of the optimal individual financial decision making related to retirement. Two important risk factors are longevity risk and inflation. The financial instruments that can be exploited to hedge these risk are annuities, see e.g. Poterba and Wise (1998), and real bonds, see e.g. Campbell and Viceira (2001). Participation in a pension scheme usually does provide coverage against longevity risk and aims for inflation indexation and will therefore usually have substantial added value for a naive investor.[5]

Life expectancies are different among people, which have a welfare effect on individuals participating in a mandatory pension plan. Brown (2002, 2003) documented unequal expected lifetimes for groups with different characteristics. Women live longer than men, and there are significant differences in life expectancies along racial/ethnic lines. Brown (2003) documented 6 years longer life expectancy for women than for men at the age of 22 in the total US population. However these differences vary along ethnic lines. 22-year-old white men live 6.5 years longer in expectation than black, and the difference is 4 years in favor of white women. Life expectancy varies with education. White men at the age of 22 with college education live 5.2 years longer in expectation than white men with less than high school education. This difference is 7.6 years for black men, 3 years for white women and 4.4 years for black women. The differences mentioned above are still present, but slightly smaller for people with higher attained age. For instance, the expected lifetime is 3.7 years longer for women than for men with

[5] This is not only true for a naive investor, but in many countries the markets for both annuities and inflation-linked securities are underdeveloped. Pension funds therefore complete the market with respect to these two risk factors.

the age of 67. White men at the age of 67 have 2.1 years longer life expectancy than white women, etc.

Differences in life expectancies are also present in Europe. Kunst (1997) found the effect of different educational levels on life expectancy in several European countries. Huisman et al. (2004) also documented mortality differences among cohorts with different educational levels in 11 European populations. A recent report by Herten et al. (2002) documents similar findings to Brown (2003) in the Netherlands. On the basis of a social economic survey between 1995 and 1999, women at the age of 20 are expected to live 5.4 years longer than men, while this difference between women and men slightly decreases to 4.7 for people who attained the age of 65. The difference in expected lifetime is present among cohorts with different educational level. A 20-year-old highly educated man is expected to live 5 years longer, than a man at the same age with the lowest education. This difference shrinks to 3.7 years as soon as he reaches the age of 65. A 20-year-old woman with high education lives 2.6 years longer in expectation than a woman with the lowest education, and this difference becomes 2.1 years as a woman gets 65 years old.

Differences in life expectations induce wealth transfers among different cohorts, distinguished along gender, educational level or ethnic lines. If cohorts with different characteristics are pooled and participate in a collective pension or annuity plan that does not take into account cohort-specific differences, people with worse survival prospect subsidize groups with higher expected lifetime. This statement equally holds for pension funds setting premium or for insurance companies selling annuity products.

For instance, Brown (2002, 2003) examined the distributional implications of compulsory annuitization in the US by using the money‘s worth of annuity framework. The money‘s worth measure is the expected present value of annuity payments per dollar spent to purchase the annuity. If annuities are qualified (payment received each month from a qualified annuity is taxable as income) annuity rates are generally unisex, which implies that the monthly annuity payments are constrained to be the same (uniform) for all individuals. Brown (2002, 2003) report the money‘s worth of the uniform annuities for individuals with the age of 67, by taking into account cohort-specific (gender, educational, race) survival characteristics. In expectation men pay 6.6% more and women pay 5.6% less than the present value of the nominal annuity they are expected to receive. Black men pay 12.9%, however white men pay 5.9% more than the fair value of the annuity. Black women pay 1.1%, while white women pay 6.3% less. Moreover, highly (college) educated white men pay only 2% more and low educated white men pay 12.5% more than the market value of their nominal annuity. On the other hand, highly educated white women pay 7.9% less, while low educated white women pay 3% less than the market value of their annuity they are expected to receive. Similar patterns can be observed while looking at the effect of educational differences for black men and women.

Brown (2002, 2003) calculated the money's worth of annuity values for real annuities as well. Cohorts which gain in the nominal plan will also gain with the real annuity. Similarly, the same is true for losses. In the whole population, the losses suffered by men are 8.7%, and the gains for women are 7.1% in real terms. The losses and gains are higher than in nominal plans, and this statement holds for all race- or education-specific cohorts in almost all cases.

Feldstein and Liebman (2002) calculated the net present value of the lifetime participation for different cohorts in the US population in a funded pension system. Participants pay 9% payroll taxes to a personal retirement account (PRA) which earns 5.5% return and the balance is fully annuitized when the individual reaches retirement. PRA annuities are calculated by using a single uniform unisex mortality table; age-, sex-, race-, and education-specific differences are ignored. The results are sensitive to the choice of the discount rate which is used to calculate the net present values, however, most of the conclusions are robust to its size. The net present value of the lifetime contribution is higher for women than for men, and white people benefit more than black. The results related to differences in the educational level depend on the size of the discount rate. If the discount rate is 1% or 3%, then higher education groups benefit more than cohorts with low education. However, if the discount rate is assumed to be 5%, then cohorts with the highest and lowest education benefit almost the same, however the group with middle level education benefits the most.

Many of the results discussed in this chapter can also be applied to pricing annuity products that are offered to individuals. The potential important additional complication there is that of adverse selection. A well-known stylized fact in the annuity literature is that those that choose to buy an annuity have a life expectancy that exceeds that of the population at large (see Mitchell et al., 1999; Finkelstein and Poterba, 2002, 2004). In this chapter we ignore potential information asymmetries.

5.3 The money's worth of participation in collective pension schemes

In this section we consider the economic value of participation in a collective pension fund which offers purely nominal or purely real pension benefits to the participants.

The participation in a pension fund is compulsory, i.e. all employees have to participate collectively in a fund. We consider three types of funds. The nominal fund offers guaranteed (DB) nominal benefits after retirement. The real fund offers benefits that do not deteriorate in real terms, i.e. are protected against inflation. Subsequently we will also consider conditionally indexed funds which offer inflation indexation if the funding ratio is sufficiently high. We restrict our attention to defined benefit pension systems with uniform pricing, i.e. all participants of the pension plan pay the same fraction of the salary. This means the contribution rate is set by the fund uniformly among participants in percentage terms, and each member contributes the same percentage of his/her yearly salary irrespective of the individual characteristics.

The current institutional setting in the Netherlands is such that pension rights are built up for every year worked. For a one year participation in the average-wage[6] pension system, the employee earns the right to receive 1.75% of the current yearly wage after retirement[7]. The amount of pension benefit is maximized at 70% of the average wage over the career, and possibly corrected for inflation.

Differences in survival rates, income profiles and in particular deferred time of the annuity imply differences in the value of participation among cohorts. We distinguish the participants along age, gender and educational level. We assume that people can contribute to the fund from the age of 25 till age 64. In addition, we distinguish 5 educational groups[8], such as people only with basic, lower secondary, higher secondary, high education, and education level which reflects the population average.

Survival probabilities for different socioeconomic groups in the Netherlands are not easily available. In Appendix 5.A we explain how estimated survival probabilities per group have been constructed using Dutch data for the population at large and Belgian data on socioeconomic survival probabilities.

[6]In case of final wage, there is a phenomenon called back service, which means that additional money has to be contributed if people have steep career patterns. If the average wage scheme is used, it can be shown that back service is no longer important.

[7]In average wage systems, the participant receives 70% of the average wage after 40 years of participation. This is equivalent with the fact, that one year participation yields the right to receive 1.75% of the current wage after retirement.

[8]Basic education means the primary level education, which is 8 years of school. The second educational group, the lower level of secondary education is defined as the level of education reached after three years of primary education. The third group with the higher level of secondary education consists of people who have 6 years education after primary school. The fourth group has higher or university degree. In addition, we create a group which portrays the average education of the total population of the Netherlands.

There are sizable differences in expected lifetime[9] among cohorts with different educational levels. First, we calculated these differences by assuming that survival probabilities do not improve in the future, and the calculations have been made on the basis of the latest Dutch life table observed in 2003. However, due to improvements in health care or in living standards etc., survival probabilities may change over time. A parsimonious model to capture the dynamics of the survival probabilities is the Lee-Carter (LC hereafter) model as introduced by Lee and Carter (1992). The details on the model and estimation are provided in Appendix 5.B. Therefore, as an alternative to the no improvement in survival assumption, we accommodate the projected improvement in survival probabilities and recalculated the differences in expected lifetime between cohorts. Table 5.1 shows the educational-, age-, and gender-specific expected lifetime and probability of survival with constant mortality and with mortality improvement.

If survival rates are assumed to be constant, a 25-year-old man with high education is expected to live 4.8 years longer than a man with basic education, while this difference is 3.2 years for a woman with the age of 25. The differences in life expectations between high and basic educational groups decrease to 2.9 for men and 2.3 for women at the age of 64.

The model with time-varying survival rates predicts further increases of life expectancy.[10] The projected life expectancy at age 25 of a man with average education increases with 2.3 years if the assumption of constant mortality rates is dropped and similar differences hold for other educational levels as well. The corresponding difference for women is 4. The differences decrease to 0.4 years in the case of men, and to 0.9 in the case of women at the age of 64. Since the methodology we use (see Appendix 5.A on estimating socioeconomic life tables for the Netherlands) to calculate educational-specific survival rates makes sure that the relative differences in gender-specific life expectations between educational cohorts do not change, or at least do not decrease in the future (for details, see Pappas et al., 1993; Preston and Taubman, 1994; Mackenbach et al., 2003), the educational-specific differences in life expectancy with mortality improvement are similar to the results based on the constant mortality assumption.

The present value of a nominal (real) annuity contract depends on mortality rates

[9]Survival data for the Netherlands is downloaded from the Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 01.12.2004).

[10]Note that life expectancy does not increase monotonically with age if the projected mortality improvements are incorporated. This is due to the fact that survival rates are expected to drop considerably which at young ages dominates the effect that people have already reached a certain age.

	Constant Mortality						Projected Mortality Improvement					
	men			women			men			women		
Age	Exp. Life	Prob of surviving age 65	Prob of survivin g age 80	Exp. Life	Prob of surviving age 65	Prob of survivin g age 80	Exp. Life	Prob of surviving age 65	Prob of surviving age 80	Exp. Life	Prob of surviving age 65	Prob of surviving age 80
	Average education											
25	76.6	0.857	0.469	81.1	0.903	0.648	78.9	0.880	0.543	85.1	0.930	0.759
35	76.9	0.863	0.472	81.3	0.907	0.650	78.6	0.878	0.530	84.4	0.925	0.738
45	77.4	0.875	0.478	81.7	0.916	0.657	78.6	0.883	0.520	84.0	0.926	0.722
55	78.4	0.907	0.496	82.6	0.942	0.675	79.2	0.909	0.522	84.1	0.945	0.717
64	80.1	0.986	0.539	83.8	0.992	0.711	80.5	0.986	0.553	84.7	0.992	0.733
	High education											
25	78.7	0.888	0.538	82.0	0.912	0.672	80.9	0.906	0.607	86.0	0.937	0.776
35	78.8	0.892	0.540	82.1	0.915	0.674	80.6	0.904	0.594	85.3	0.932	0.757
45	79.2	0.900	0.545	82.5	0.923	0.680	80.4	0.906	0.584	84.8	0.932	0.741
55	80.0	0.925	0.560	83.3	0.947	0.697	80.8	0.927	0.585	84.9	0.950	0.736
64	81.4	0.988	0.598	84.5	0.992	0.731	81.9	0.988	0.612	85.4	0.992	0.751
	Higher Sec. education											
25	76.4	0.854	0.462	81.2	0.904	0.653	78.6	0.877	0.537	85.3	0.931	0.762
35	76.7	0.859	0.465	81.4	0.908	0.656	78.4	0.875	0.523	84.6	0.926	0.743
45	77.2	0.871	0.471	81.8	0.917	0.662	78.4	0.880	0.513	84.2	0.927	0.726
55	78.3	0.904	0.489	82.7	0.943	0.681	79.0	0.907	0.516	84.3	0.946	0.722
64	79.9	0.985	0.533	84.0	0.992	0.716	80.4	0.985	0.547	84.9	0.992	0.738
	Lower Sec.education											
25	75.7	0.842	0.438	80.8	0.900	0.643	77.9	0.867	0.516	84.8	0.927	0.755
35	76.1	0.850	0.442	81.0	0.904	0.646	77.7	0.866	0.502	84.2	0.923	0.735
45	76.7	0.863	0.449	81.5	0.914	0.653	77.8	0.872	0.493	83.8	0.924	0.719
55	77.8	0.899	0.468	82.4	0.942	0.673	78.5	0.902	0.496	83.9	0.945	0.715
64	79.5	0.985	0.512	83.7	0.992	0.709	79.9	0.985	0.527	84.6	0.992	0.731
	Low education											
25	73.9	0.805	0.375	78.8	0.878	0.573	76.1	0.835	0.456	82.9	0.911	0.702
35	74.5	0.816	0.380	79.2	0.884	0.576	76.1	0.835	0.442	82.3	0.907	0.680
45	75.2	0.833	0.388	79.7	0.896	0.584	76.3	0.844	0.433	82.0	0.909	0.660
55	76.5	0.877	0.408	80.8	0.928	0.605	77.2	0.880	0.437	82.2	0.932	0.654
64	78.5	0.981	0.457	82.2	0.990	0.645	78.9	0.981	0.473	83.0	0.990	0.671

Table 5.1: Educational-specific expected lifetime and probability of survival for selected age groups. The table gives the educational-, age-, and gender-specific expected lifetime and probabilities with the constant mortality (at the level estimated for 2003), and the time-varying future mortality assumptions.

as well as on the nominal (real) term structure. If tax considerations are ignored, the present value $V^i_{x,t}$ at time t for an individual i with age x of an A dollar nominal annuity as of the age of 65 can be written as (compare e.g. Mitchell et al., 1999):

$$V^i_{x,t} = \sum_{s=65-x}^{\infty} \mathbb{E}_t \left(p^i_{x,t}\right) \frac{A}{\left(1+R^{(s)}_t\right)^s}, \tag{5.1}$$

where ${}_sp^i_{x,t}$ is the probability at time t that person i at the age of x is going to live at least for another s years, and $R^{(n)}_t$ is the nominal interest rate at time t for payment in n periods from now. The same expression applies for the value of a real annuity $V^{i,R}_{x,t}$ if the nominal interest rate is replaced by the corresponding real interest rate $R^{R(n)}_t$. In our case A denotes the right to receive 1.75% of the current yearly wage after retirement either in a nominal or real indexed scheme, and $\mathbb{E}_t\left({}_sp^i_{x,t}\right)$ can be calculated by using either the no improvement or improvement assumption in future survival rates. For a detailed derivation of (5.1), see Appendix 5.B and 5.C, where it is moreover shown that this expression can be extended to the case of conditionally indexed schemes.

The real ($R_t^{R(n)}$) and nominal ($R_t^{(s)}$) term structures used to calculate the present value of the one year participation in a pension fund are presented on Figure 5.1, which corresponds to a nominal 10-year rate of 4.2% and inflation rate of 1.2% p.a. in January 1, 2004, in the Netherlands.

The money's worth of a fixed (in nominal or real terms) annuity in a collective scheme is presented in Tables 5.2 and 5.3. These tables generalize Table 2 in Brown (2002) by providing evidence not only of the annuity value at retirement, but also of the money's worth in the accumulation phase of the life-cycle, thereby adding the age dimension. The figures in the tables show the present value of one year participation for age-, gender-, and educational-specific cohorts.

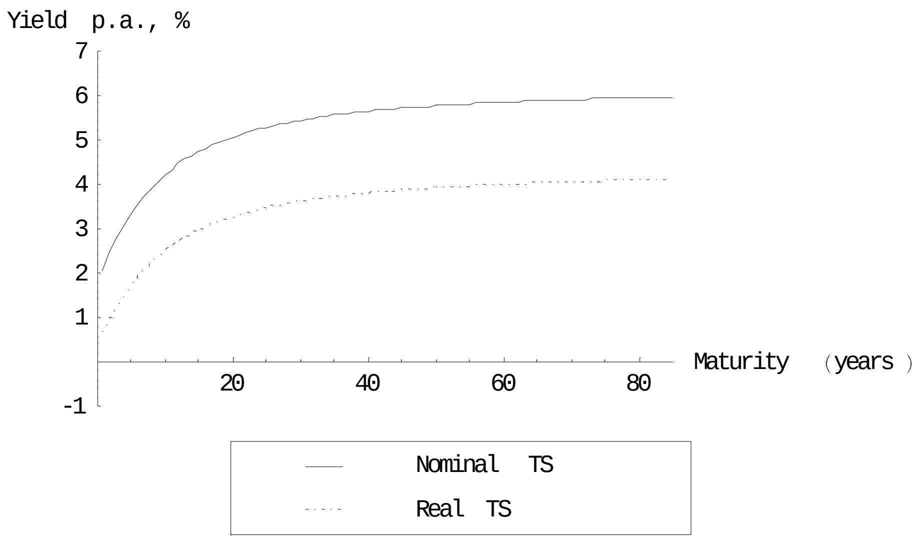

Figure 5.1: Nominal and real term structures that are used to determine the money's worth of participation in a collective scheme, January 1, 2004. The figure illustrates the nominal and the real term structure of interest rates, which corresponds to a nominal 10-year rate of 4.2% and inflation rate of 1.2% p.a. in January 1, 2004, in the Netherlands.

Table 5.2 gives the money's worth of participation if survival rates are constant and the term structure of interest rates in Figure 5.1 is used. A 35-year-old woman with

Constant mortality			Low	Lower Sec.	Higher Sec.	High	Average
Nominal Annuity	Men	At 25 yrs	1.37%	1.49%	1.54%	1.68%	1.55%
		At 35 yrs	2.59%	2.81%	2.89%	3.15%	2.91%
		At 45 yrs	4.92%	5.32%	5.45%	5.91%	5.50%
		At 55 yrs	9.52%	10.17%	10.40%	11.18%	10.47%
		At 64 yrs	17.39%	18.23%	18.54%	19.60%	18.65%
	Women	At 25 yrs	1.71%	1.83%	1.85%	1.89%	1.84%
		At 35 yrs	3.21%	3.43%	3.47%	3.54%	3.45%
		At 45 yrs	6.06%	6.46%	6.52%	6.65%	6.49%
		At 55 yrs	11.55%	12.24%	12.35%	12.55%	12.28%
		At 64 yrs	20.22%	21.23%	21.40%	21.68%	21.30%
Real Annuity	Men	At 25 yrs	3.23%	3.54%	3.65%	4.03%	3.69%
		At 35 yrs	5.06%	5.53%	5.69%	6.26%	5.75%
		At 45 yrs	7.98%	8.67%	8.91%	9.75%	8.99%
		At 55 yrs	12.82%	13.77%	14.11%	15.29%	14.23%
		At 64 yrs	19.94%	21.02%	21.44%	22.85%	21.58%
	Women	At 25 yrs	4.11%	4.44%	4.50%	4.61%	4.47%
		At 35 yrs	6.41%	6.90%	6.99%	7.15%	6.94%
		At 45 yrs	10.03%	10.77%	10.90%	11.13%	10.83%
		At 55 yrs	15.86%	16.95%	17.12%	17.45%	17.02%
		At 64 yrs	23.66%	25.04%	25.28%	25.68%	25.14%

Table 5.2: The present value of participation in collective pension funds with no mortality improvement. The table gives the educational-, age-, and gender-specific money's worth of participation as a percentage of the yearly salary in a nominal and real pension scheme if survival rates are constant over time at the level estimated for 2003.

low education earns 3.21% of the current yearly salary if she participates in a nominal pension scheme, while she earns 6.41% of the current salary in a real pension plan, twice as much as in the nominal case.

The numerical results on the money's worth of one year participation in a nominal fund with different characteristics of people can be summarized as follows. For a given age, the money's worth of a single year participation is decreasing as educational level decreases. For instance, a 25-year-old man earns 1.68% of the annual salary if he is highly educated, and 1.37%, if he attained basic education only. A 25-year-old highly educated man earns 22.6% more pension right than a man with basic education at the same age, which is due to different survival prospects. The difference between the money's worth of participation due to different level of education shrinks to 12.7% for men at the age of 64. The differences in the money's worth measure among the highest and lowest educated groups with the same age are also present for women, however they are somewhat smaller. A 25-year-old woman with the highest education earns 10.5% more than someone with the lowest education, while the difference decreases to 7.2% at the age of 64. Due to different gender-specific survival rates a 25-year-old man with low education is expected to earn 1.37% of the yearly salary, a woman with the same age with the same level of education earns 1.71% of the salary, 24.8% more.

Gender- and education-specific survival differences at a given age are reflected in real pension plans as well. The pattern of differences is similar, but the differences in percentage terms are higher for real pension plans, that are analyzed in the lower part of Table 5.2. This is due the fact, that real interest rates are smaller than the nominal ones, therefore the differences in real pension rights due to different survival characteristics of people with the same age are less affected by the effect of discounting than in the case of nominal pension schemes. The difference between the present value of participation earned by low and highly educated men is 24.8% at the age of 25 and it is 14.6% with the age of 64. For women the corresponding numbers are 12.2% and 8.5%.

Table 5.2 clearly shows that the age of the participants has a very important role in determining the present value of nominal annuity earned by a one year participation in the fund. In the nominal scheme, the present value of participation for a woman with low education is 20.22% of her yearly salary at the age of 64, and it is 1.71% of the salary at the age of 25, which is 91.5% less. This is caused by two effects. One is the probability of death, and the other one is the time value of money. Table 5.1 clearly shows the uncertainty effect in survival. A 25-year-old woman with low education has an 88.7% (see Table 5.1, column 11, 0.878/0.99) probability to survive till the age of 64. This makes the present value of the annuity decrease by 11.3%. However, the discounting makes the present value of annuity decrease further by another 80.2%. The effect of discounting dominates the differences in the money's worth of annuities among groups with different ages. If the term structure shifts downwards, the effect of discounting is obviously less strong. Consequently, the differences in the present value of participation in the real pension plan is smaller (the corresponding number for 91.5% in nominal the plan becomes 82.6% in the real plan) due to the lower real yields.

In Table 5.2 we assumed that the survival probabilities as observed in 2003 will not improve further. Table 5.3 presents similar results, but assumes projected mortality improvements as discussed in Appendix 5.B and presented in Table 5.1. Table 5.3 shows the value of participation if survival probabilities are time-varying, the nominal 10-year rate is 4.2% and the 1-year inflation rate is 1.2%.

The benefits are obviously higher compared to the case with time-invariant survival probabilities, because the probability of surviving increased for all age groups. Adjustment for mortality improvement has an impact of up to 50-85 bp on the money's worth

Mortality Improvement			Low	Lower Sec.	Higher Sec.	High	Average
Nominal Annuity	Men	At 25 yrs	1.52%	1.63%	1.67%	1.81%	1.68%
		At 35 yrs	2.79%	3.00%	3.08%	3.33%	3.10%
		At 45 yrs	5.18%	5.56%	5.69%	6.15%	5.74%
		At 55 yrs	9.80%	10.44%	10.67%	11.44%	10.74%
		At 64 yrs	17.62%	18.46%	18.78%	19.85%	18.88%
	Women	At 25 yrs	1.94%	2.04%	2.06%	2.10%	2.05%
		At 35 yrs	3.55%	3.74%	3.78%	3.84%	3.76%
		At 45 yrs	6.51%	6.87%	6.94%	7.05%	6.90%
		At 55 yrs	12.06%	12.72%	12.83%	13.03%	12.77%
		At 64 yrs	20.69%	21.68%	21.85%	22.13%	21.75%
Real Annuity	Men	At 25 yrs	3.62%	3.91%	4.02%	4.38%	4.05%
		At 35 yrs	5.52%	5.96%	6.12%	6.68%	6.18%
		At 45 yrs	8.46%	9.13%	9.38%	10.21%	9.46%
		At 55 yrs	13.26%	14.21%	14.55%	15.74%	14.67%
		At 64 yrs	20.28%	21.36%	21.79%	23.22%	21.92%
	Women	At 25 yrs	4.77%	5.05%	5.10%	5.20%	5.08%
		At 35 yrs	7.19%	7.64%	7.72%	7.87%	7.68%
		At 45 yrs	10.90%	11.60%	11.72%	11.95%	11.66%
		At 55 yrs	16.71%	17.77%	17.95%	18.27%	17.84%
		At 64 yrs	24.35%	25.72%	25.97%	26.38%	25.82%

Table 5.3: The present value of participation in collective pension funds, mortality improvement. The table gives the educational-, age-, and gender-specific money's worth of participation as a percentage of the yearly salary in a nominal and real pension scheme if survival rates are allowed to be time-varying.

for some age groups, depending on the scheme. In a nominal pension scheme, a 25-year-old highly educated man earns 19.1% more pension right than a man with basic education at the same age, if he participates in the fund for a year. This difference shrinks to 12.7% for men at the age of 64. A 25-year-old woman with the highest education earns 8.2% more than someone with the lowest education, while the difference decreases to 7.0% at the age of 64. A 25-year-old woman with low education earns 27.6% more than a man of the same age. The money's worth of one year participation is 90.6% lower for a 25-year-old low educated woman, than for a woman with the same characteristics with the age of 64. Similar conclusions hold for real pension schemes.

In pension funds with uniform contribution rates the cost-effective contribution rate is set to cover the market value of the rights assigned to the participants.[11] The deviation between this cost-effective rate for the fund as a whole and the percentage of the wage that reflects the money's worth of the annuity indicates the cost solidarity, that is imposed by the fund; i.e it indicates whether an individual is a net contributor or

[11] This contribution rate varies typically from 12.5% to 17.5% of the yearly salary in the Netherlands.

beneficiary of the scheme in the year under consideration.[12] Our analysis quantifies the solidarities in the typical Dutch pension deal (referred to e.g. by Kune, 2005): from young to old, from men to women, from lower educated to higher educated.

If the costs of switching to another fund are low, and the additional assumptions made in the analysis apply (in particular the assumption that agents unwind the positions imposed by the fund so that only market values are relevant), some groups are better off if they leave the compulsory fund. The uniform premium creates an incentive for young cohorts to avoid the compulsory scheme, e.g. by reducing their labor market participation, by moving abroad or by becoming self-employed. This finding is valid both for nominal and real schemes. Leaving the age effect aside, individuals, whose money's worth of the imposed annuity is lower than the uniform premium that is charged, have an incentive to buy a tailor-made annuity on the private insurance market if such annuities are offered. If opting out of the compulsory system is feasible at relatively low cost, the sustainability of the compulsory system of course becomes questionable.

5.4 The money's worth in conditionally indexed collective pension schemes

Pension plans in many countries are typically neither nominal nor real in nature, instead, they are hybrid constructs. Pension funds typically offer inflation indexation of accumulated pension rights if the current state of the fund is good. However the rules of indexation are often not specified explicitly in the contract.

In the Netherlands many pension funds have recently made more or less explicit their indexation promise. If the nominal funding ratio drops below a certain threshold no indexation is granted and the premium is increased. If the nominal funding ratio is above a certain threshold, full indexation is granted and contributions are decreased. In between the thresholds, partial indexation is granted and contributions decrease as the funding ratio increases.

[12] If the uniform contribution rate is 12.5% in the nominal plan and mortality improvement is taken into account (Table 5.3) for instance, then a highly educated 25-year-old woman is a net contributor. She pays 12.5% of her salary as a contribution and the present value of her yearly participation is 2.1% of the yearly salary, implying a 10.4% net contribution. However, a woman with the same education level but with the age of 64 is a net beneficiary. She pays 12.5% of the salary and receives 22.13% in exchange for the participation; she benefits 9.63% of her yearly salary. A similar cost-benefit analysis can be carried out based on a given uniform contribution rate for funds in Table 5.2 or in Table 5.4.

In the sequel we determine the fair value of such a contract, and indicate how it depends on the funding ratio and the asset mix that is selected. We assume that pension rights are fully indexed against inflation if the nominal funding ratio is larger than 1.36, they are partially indexed if the nominal funding ratio is between 1.05 and 1.36, and no indexation occurs below the 1.05 level.

The fund we consider is a large[13] closed-end fund[14] where no premium inflow takes place and no new benefits are built up in the future on the basis of the discontinuity perspective. The liabilities of the fund are used to pay future pension liabilities, and the pension benefits are indexed according to the indexation rules discussed in the previous paragraph. We assume a specific linearly decreasing expected nominal liability stream with duration of 13.4 years, illustrated on Figure 5.2:

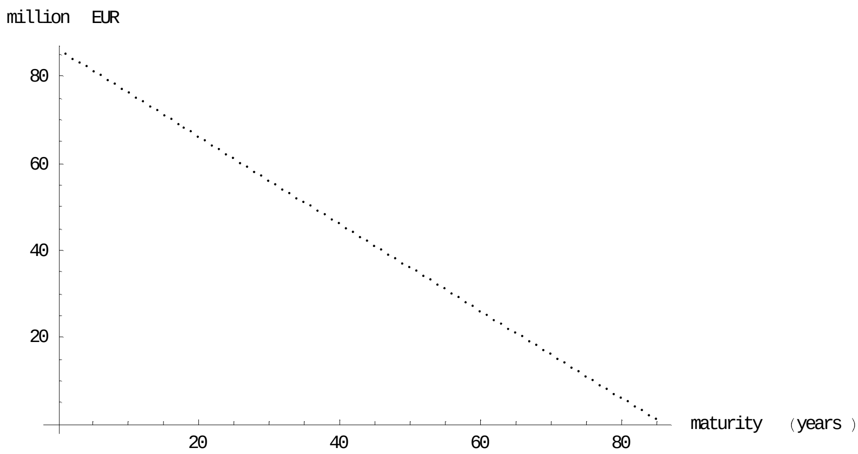

Figure 5.2: Liability stream of the collective fund. The figure shows the expected liability stream of a large pension fund up to the maturity of 85 years, where no premium inflow takes place and no new benefits are built up in the future on the basis of the discontinuity perspective.

The variable annuities that are offered by participation in a conditionally indexed scheme can be priced by using the pricing kernel. The pricing kernel is a particular random variable so that the price of any asset at time t satisfies $P(t) = E(P(t +$

[13]Micro-longevity risk, which results from nonsystematic deviations from an individuals expected remaining lifetime is negligible in the case of a large fund.

[14]Alternatively, we could set up a running fund and policy rules are applied for setting the size of the contribution in each year. This would yield a more realistic liability stream but would complicate the analysis.

$1)M(t+1)|F(t))$[15]. This implies that the risk premium of any asset is determined by its covariation with the pricing kernel. If markets are complete, i.e. every contingent claim can be replicated by a self-financing portfolio, this pricing kernel is uniquely given. In this case, the price obtained via valuation using the pricing kernel can be shown to be equivalent to the cost required to set up the replicating portfolio. In incomplete markets, however, we cannot identify all risk premia on the basis of the traded assets. In our context, market incompleteness is for instance caused by macro-longevity risk[16], if no annuities are traded, and inflation, if no inflation-linked securities are available. In such situations, one has to make assumptions regarding the pricing kernel specification. We assume that neither longevity risk nor inflation risk is priced.[17] For a detailed specification of the pricing kernel and a discussion regarding the assumptions made, we refer to Appendix 5.C and to Nijman and Koijen (2006).

Table 5.4 shows the value of the conditionally indexed rights for constant mortality at the level estimated for 2003, and for time-varying mortality rates. The value of the conditionally indexed pension rights are bounded by the value of a nominal and a real plan. If the current funding ratio is low, the likelihood of indexation is low and as a consequence, the conditionally indexed scheme resembles a nominal scheme. If the current funding ratio is high, the pension scheme is highly comparable to a real pension scheme.

Note that the asset mix influences the money's worth of participation in a conditionally indexed scheme. Since the nominal rights are guaranteed, the money's worth of participation in a scheme with low funding ratio increases if more risk is taken. If the fraction of assets invested into stocks reaches a certain threshold, the money's worth starts decreasing for some age groups. Likewise, a large fraction invested in stocks reduces the value of participation if the funding ratio is high. The reason of the decline in the money's worth if the risk increases is the fact that the sponsor gets the up-side above the full indexation.

For the constant mortality case, if the initial nominal funding ratio is 1, a 64-year-old

[15]The existence of a pricing kernel is ensured when the financial market is free of arbitrage, which we will assume throughout. For more details on pricing kernels, we refer to Campbell et al. (1997) and Cochrane (2001).

[16]Macro-longevity risk results from the fact that survival probabilities change over time.

[17]There is an empirical evidence that investors demand risk-premium for holding inflation-sensitive assets (see the UK nominal and inflation-linked gilt market for instance in Evans, 1998), however, for simplicity we assume that inflation risk is not priced, because we do not observe inflation-linked instruments in the Netherlands.

		FR=1			FR=1.4			Nominal Plan	Real Plan
		0% Stocks	50% Stocks	100% Stocks	0% Stocks	50% Stocks	100% Stocks		
Constant Mortality									
Men	At 25 yrs	1.74%	1.98%	1.91%	3.28%	2.63%	2.21%	1.55%	3.69%
	At 35 yrs	3.08%	3.50%	3.44%	5.21%	4.41%	3.89%	2.91%	5.75%
	At 45 yrs	5.61%	6.24%	6.23%	8.37%	7.46%	6.85%	5.50%	8.99%
	At 55 yrs	10.53%	11.29%	11.37%	13.66%	12.75%	12.15%	10.47%	14.23%
	At 64 yrs	18.66%	19.26%	19.40%	21.24%	20.57%	20.13%	18.65%	21.58%
Women	At 25 yrs	2.08%	2.36%	2.28%	3.95%	3.16%	2.65%	1.84%	4.47%
	At 35 yrs	3.67%	4.16%	4.09%	6.26%	5.28%	4.64%	3.45%	6.94%
	At 45 yrs	6.65%	7.40%	7.38%	10.02%	8.89%	8.16%	6.49%	10.83%
	At 55 yrs	12.37%	13.30%	13.39%	16.25%	15.11%	14.37%	12.28%	17.02%
	At 64 yrs	21.33%	22.11%	22.27%	24.65%	23.77%	23.20%	21.30%	25.14%
Mortality Improvement									
Men	At 25 yrs	1.90%	2.15%	2.08%	3.61%	2.89%	2.41%	1.68%	4.05%
	At 35 yrs	3.29%	3.72%	3.67%	5.59%	4.74%	4.14%	3.10%	6.18%
	At 45 yrs	5.86%	6.52%	6.52%	8.78%	7.82%	7.16%	5.74%	9.46%
	At 55 yrs	10.80%	11.58%	11.67%	14.04%	13.10%	12.47%	10.74%	14.67%
	At 64 yrs	18.90%	19.51%	19.67%	21.55%	20.86%	20.41%	18.88%	21.92%
Women	At 25 yrs	2.35%	2.65%	2.56%	4.51%	3.57%	2.98%	2.05%	5.08%
	At 35 yrs	4.03%	4.56%	4.48%	6.94%	5.81%	5.09%	3.76%	7.68%
	At 45 yrs	7.10%	7.90%	7.89%	10.79%	9.54%	8.71%	6.90%	11.66%
	At 55 yrs	12.88%	13.86%	13.96%	17.03%	15.79%	14.97%	12.77%	17.84%
	At 64 yrs	21.80%	22.61%	22.78%	25.30%	24.36%	23.74%	21.75%	25.82%

Table 5.4: The value of conditionally indexed rights. The table gives the value of the age- and gender-specific conditionally indexed rights as a percentage of the yearly salary for an annuity population with an average education. We consider a fund with a starting nominal funding ratio of 1, and alternatively, we consider another fund with identical characteristics, except, that the starting nominal funding ratio of the latter is 1.4. We allow for alternative investment strategies. The assets are invested into 10-year nominal bonds and stocks. The results of the conditionally indexed schemes are compared to the purely nominal and real plans. In the upper part of the table survival probabilities are constant over time at the level estimated for 2003, while in the lower part survival probabilities are time-varying.

woman earns 21.33% of the salary if 100% of the assets are invested into 10-year nominal bonds, while the present value of participation increases to 22.11% with a 50% 10-year nominal bonds and 50% stock asset mix, and it further increases to 22.27% with an asset mix of 100% stocks. If the initial funding ratio is 1.4, the increase in the risky assets in the asset mix yields a lower value for the value of participation, because the probability of ending up in the bad state is higher with more volatile stock investments. The pattern is similar if mortality rates are time-varying, and the present value of participation is higher due to the improvement in expected lifetime.

The main conclusion is that all members, regardless of the age of the participants, have an incentive[18] to opt out of a fund with a low funding ratio to a fund with a higher funding ratio. The value of participation is lower and the cost of participation is the same or higher than in a fund with a higher funding ratio. In the case of time-varying mortality, the value of participation is 2.15% of the yearly salary for a 25-year-old man if he participates in a fund with the initial funding ratio of 1 and the asset mix of 50%-50%. This person has an incentive to change pension fund, because the identical fund with a higher funding ratio is more appealing. The value of the one year participation increases to 2.89% of the yearly salary. A 64-year-old man with earns 19.51% of the yearly salary in a fund with funding ratio of 1 and 50% of stocks. However, if the funding ratio increases to 1.4 and all the other characteristics of the fund remain the same, then this person earns 20.86%.

Besides the indexation quality, the other motives generated by the gender-, education- and age-specific survival differentials still play an important role for opting out of the collective pension plan, either individually or collectively. However, we do not want to replicate those arguments again in this section.

5.5 Conclusions

The money's worth of participation in a collective fund is different among age groups and socioeconomic groups. Young participants have an incentive to opt out if uniform pricing over age groups is applied. Similar but much smaller differences occur between socioeconomic groups and male/female participants. Generally, the money's worth of participation for lower educated is lower than for higher educated cohorts, and the value of participation for men is less than for women. Young, lower educated and males have the incentive to leave the collective fund, and to switch to another job, sector, where the characteristics of the participants of the new fund are closer to the characteristics of the people with the incentive to opt out. Alternatively, they can reduce their labor supply or try to obtain access to insurance products that are priced on the basis of their individual characteristics.

Indexation quality is another factor that affects money's worth. The money's worth of participation in a fund with a low funding ratio is less, therefore participants of a fund with a low funding ratio might have an incentive to switch to a fund with a higher funding ratio. It should be noted that switching from fund to another may have implications

[18]Transfer value of pension rights is often calculated with actuarial valuation, which might distort the incentives induced by the indexation quality and the asset mix.

for the rights built up. If so, then the incentives provided should be balanced with the switching costs. We find that especially young individuals might have the incentive to opt out, and this is exactly the group for which the rights built up are the lowest.

If people start opting out on large scale, a pension scheme is not sustainable. Consequently, the arguments underlying uniform pricing should be carefully reconsidered.

Note that a transition from uniform pricing in collective pension schemes to pricing on the basis of market value conditional on age (and possibly also on other individual characteristics) generates a substantial transitional problem, not unlike that of a transition from pay-as-you-go (PAYG) to funded systems. Young generations that have received the implicit promise that their money's worth would be more than their contribution during the last part of their working life, will have to be compensated if uniform pricing over age groups would be abolished.

In this chapter we made a number of strong assumptions. Only old-age pension has been considered and part of the money's worth differentials that have been identified can be compensated by the partner pension arrangements that are usually also included in actual pension products. Moreover we made the strong assumption that only the market value of what is received is relevant because agents can and will unwind all product features that are imposed by the pension fund. Subsequent analysis will have to consider the question to what extent these assumptions dominate the analysis.

5.A Socioeconomic life tables

This section addresses the methodology that has been used to construct the educational-specific cohort life tables. The main complication is that these data are not publicly available for the Netherlands. Deboosere and Gadeyne (2002) calculated educational-specific cohort life tables for Belgium for the period of 1991-1996. We use their results to estimate survival probabilities per socioeconomic group for the Netherlands. Deboosere and Gadeyne (2002) distinguish 4 educational levels, namely low education (L), lower secondary (SL), higher secondary (SH), and high (H).

Following Brown et al. (2002), we construct educational-specific cohort life tables, assuming that relative discrepancies in mortality rates between different socioeconomic groups are constant over time. Although it is hard to verify this assumption, Pappas et al. (1993), Preston and Taubman (1994) and Mackenbach et al. (2003) document that differences in mortality rates between socioeconomic groups are not shrinking in the late 20th century, instead, there may have been widening. If the latter is true, then we underestimate the differences between educational groups. Therefore, our assumption, if not satisfied, seems to result in conservative estimates of the differences between different educational groups. Secondly, we assume that the differences between socioeconomic groups in Belgium provide a reasonable representation of the Dutch population.

First of all, we construct the relative discrepancies from the average mortality rates for all socioeconomic groups:

$$\Delta_x^i = \frac{\hat{q}_{x,t^*}^i}{\hat{q}_{x,t^*}}, \tag{5.2}$$

on the basis of the Belgian data, where $\hat{q}_{x,t^*}^i$ indicates the 1-year mortality rate at time t^* for a person of age x that is within socioeconomic group i, $i = L, SL, SH, H$. $\hat{q}_{x,t^*}$ is the weighted average of all Belgian mortality rates at time t^*, where the weighting occurs via the number of people present in the socioeconomic group in the Netherlands:

$$\hat{q}_{x,t^*} = \frac{\sum_i N_{i,t^*} \hat{q}_{x,t^*}^i}{\sum_i N_{i,t^*}}, \tag{5.3}$$

where N_{i,t^*} indicates the number of people with age x present in socio-economic group i in the Netherlands[19]. Secondly, we apply these ratios to the Dutch population in order to calculate educational-specific mortality rates at all future points in time

$$q_{x,t}^i = q_{x,t}\Delta_x^i, \tag{5.4}$$

[19]The educational distribution of the Dutch active population for year 2002 used in the calculations were downloaded from CBS Netherlands (http://statline.cbs.nl)

where $q^i_{x,t}$ is the probability that an individual with age x at time t who is within socioeconomic group i, dies in the next year. The weighting in (5.2) is important since the composition of the Belgian population may be different than the Dutch population.

Finally, we fit cubic polynomials to smooth the ratios for different ages, and we use the smoothed ratios in order to calculate educational-specific cohort life tables for the Netherlands.

5.B Modeling survival probabilities

A crucial element in the determination of the money's worth of participation in a collective pension scheme is $\mathbb{E}_t\left(I_{(\tau<S)}(x,\tau)\right)$, see (5.30). In this section we discuss a convenient way to model mortality rates, as has been introduced by Lee and Carter (1992).

Let L denote the realization of the uncertain life table. By using the law of iterated expectations we can rewrite $E_t\left(I_{(\tau<S)}(\tau)\right)$ as follows:

$$\mathbb{E}_t\left(I_{(\tau<S)}(x,\tau)\right) = \mathbb{E}_t\left(\mathbb{E}_t\left(I_{(\tau<S)}(x,\tau) \mid L\right)\right) = \mathbb{E}_t\left({}_\tau p_{x,t}\right), \tag{5.5}$$

where ${}_\tau p_{t,x}$ is the probability that an individual with age x at time t is going to survive at least till year τ.

In the subsequent model we assume that the probability distribution of survival is uncertain. However, instead of modeling of ${}_\tau p_{x,t}$ directly, we model the time series of the log of the force of mortality $\mu_{x,t}$ [20] to calculate $\mathbb{E}_t\left(I_{(\tau<S)}(x,\tau)\right)$. In the sequel, we will assume that for any integer age x, and any time t, it holds that:

$$\mu_{x+\tau,t} = \mu_{x,t}, \qquad \text{for all } \tau \in [0,1). \tag{5.6}$$

Then, one can verify that

$${}_\tau p_{x,t} = \exp\left(-\sum_{i=t}^{t+\tau} \mu_{x+i,i}\right). \tag{5.7}$$

An important property of a model is to allow for a trend in mortality rates as has been observed historically due to improvements in medical care. A parsimonious model to capture the dynamics of the mortality rates is the Lee and Carter (LC hereafter)

[20] The force of mortality, at time t, of an individual with age x is defined as: $\mu_{x,t} = \frac{f_t(x)}{1-F_t(x)}$, where f_t (F_t) denotes the pdf (cdf) at time t of the lifetime of a newly born. For the estimation and more details, see e.g. Gerber (1997).

model. We assume that the log force of mortality is affine in a latent factor u_t, which captures the trend in mortality rates. Formally,

$$\ln \mu_{x,t} = a_x + b_x u_t + \varepsilon_{x,t}, \tag{5.8}$$

where the coefficients, a_x and b_x are age-dependent. An additional error term, $\varepsilon_{x,t}$, which is time- and age-specific, is added to capture particular age-specific influences that are not properly accounted for by the general trend. If $D_{x,t}$ denotes the number of death at time t in a cohort aged x, and $E_{x,t}$ is the number of person years, the so-called exposure, then the force of mortality can be approximated as $\mu_{x,t} \approx \frac{D_{x,t}}{E_{x,t}}$[21].

The estimation procedure of the LC model has been done in several steps. First of all, a singular value decomposition is used to retrieve an estimate of the series of the latent factor, $\hat{u}_t$. Subsequently, OLS regression are run to estimate the age-specific α_x and β_x, resulting in $\hat{\alpha}_x$ and $\hat{\beta}_x$. Once this procedure is applied, observed death numbers are generally not exactly equal to the model-based death numbers. Therefore, a correction step is made, the estimate for the latent factor at a certain point in time is adjusted so that observed number of death at time t equals the one implied by the model, i.e. $\tilde{u}_t$ solves

$$\sum_x D_{x,t} = \sum_x E_{x,t} \exp(\hat{\alpha}_x + \hat{\beta}_x \tilde{u}_t), \tag{5.9}$$

Finally, the Box-Jenkins method has been used to identify the dynamics of the latent factor u_t. The resulting specification for the latent factor driving the trend in mortality rates is

$$u_{t+1} = \mu + u_t + \eta_{t+1}, \tag{5.10}$$

with $\eta_{t+1} \overset{i.i.d.}{\sim} D(0, \sigma_\eta^2)$, i.e. the latent factor follows a random walk with drift.

In order to calculate the $\mathbb{E}_t \left(I_{(\tau<S)}(x,\tau) \right)$, we need to determine $\mu_{x,t+s}$, $s > 0$ in (5.7). We find

$$\begin{aligned} \ln \mu_{x,t+s} &= a_x + b_x u_{t+s} + \varepsilon_{x,t+s} && (5.11) \\ &= \ln \mu_{x,t} + b_x \left(u_{t+s} - u_t \right) + \varepsilon_{x,s} - \varepsilon_{x,t} && (5.12) \\ &= \ln \mu_{x,t} + b_x \left(\sum_{i=1}^{s} \eta_{t+i} + s\mu \right) + \left(\varepsilon_{x,s} - \varepsilon_{x,t} \right). && (5.13) \end{aligned}$$

Lee and Carter (1992) calculate the different sources of uncertainty in the age-specific log mortality rate, and find that the disturbance term of the latent process dominates

[21]For more details on estimating the force of mortality by the exposure and the death number, see Gerber (1997).

the error of the overall forecasted mortality rates. Lee and Carter (1992) report that in long-term forecasts about 95% of the variance is generated by innovations of the latent variable process. Therefore, we abstract from all other uncertainties. Then we find

$$\mu_{x,t+s} = \mu_{x,t} \exp\left(b_x\left(\sum_{i=1}^{s}\eta_{t+i} + s\mu\right)\right), \tag{5.14}$$

and we calculate $\mathbb{E}_t\left({}_\tau p_{x,t}\right)$ with simulation by using (5.7).

We use 100 yearly observations of number of death and exposure in the Netherlands, from 1904 till 2003[22]. We calculated force of mortality rates from the data provided by The Human Mortality Database (available at www.mortality.org or www.humanmortality.de and the data was downloaded on 01.12.2004) for 61 age groups: 25, 26, ...84, 85+. The last category denoted by 85+ refers to the average mortality rate of people with the age of 85 or older. The reason that we did not create age-groups over the age of 85 is the following. The number of people exposed to risk is relatively low in age-groups above 85 (e.g. 85-89, or 90-94 etc.) in the early 20th century. In order to get the time-series of mortality rates of elderly people calculated from sufficiently large number of observations we merged all the age-groups above year 85.

The parameter estimates of the model, which are used to forecast survival rates are as follows. As far as the latent process is concerned in (5.10), the drift term is -0.128 and the standard deviation of the disturbance term is 0.5 for women. The corresponding estimates for men are $-.095$ and 0.57 respectively.

The estimates of the age-specific sensitivity coefficients (i.e. the sensitivity of the log death rates to the change in the latent process) for women and men are illustrated on the following figures:

[22]The trends in the age-specific log mortality rates with the random walk with drift specification depend on the first and the last observations (see for instance Girosi and King, 2005b), therefore the forecasted log mortality rates are going to be sensitive to the sample period applied in the estimations. The CBS Netherlands was established in 1899, and it became the main institution responsible for collecting statistical (including population) data. Since then, data collection on births and deaths became better organized and standardized, and the data is less susceptible to measurement problems. Consequently, the sensitivity of the estimation results to the sample period suggests to use the data starting at the beginning of the 20th century.

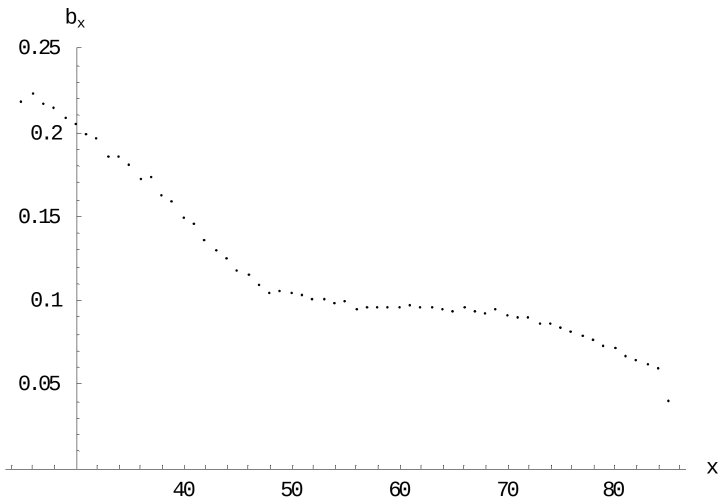

Figure 5.3: The age-specific sensitivity coefficients b_x for women. They represent the sensitivity of the age-specific log death rates to the change in the latent process.

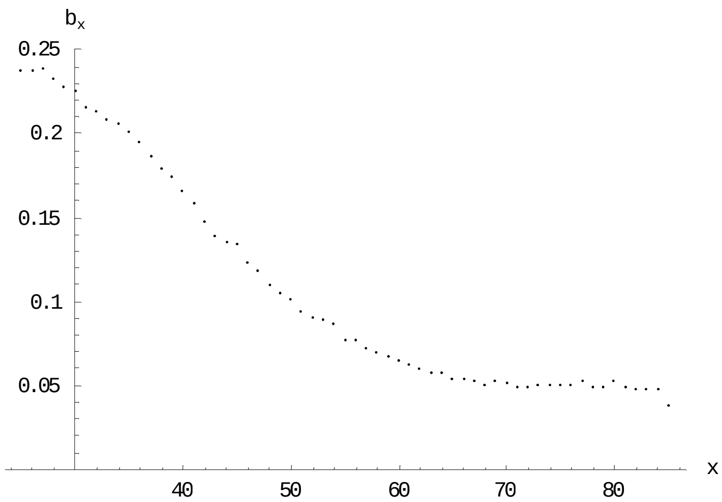

Figure 5.4: The age-specific sensitivity coefficients b_x for men. They represent the sensitivity of the age-specific log death rates to the change in the latent process.

5.C Money's worth of participation in collective pension schemes

In this appendix we outline the valuation approach that has used throughout this chapter to determine the money's worth of a single year participation. The institutional setting that has been adopted is a so-called average wage system. This implies that the average wage during the individual's working life serves as the metric to determine the pension benefit. Moreover, since inflation potentially erodes the investments of participants, pension funds often aim to provide some form of inflation indexation, which is accounted for. We assume that participation in the pension scheme implies compulsory annuitization of the accumulated benefits at the retirement date.

Formally, suppose that the individual participates l years in the collective pension scheme and receives at the beginning of each year the nominal wages $w_t, .., w_{t+l-1}$. Suppose, the individual retires at time T, with $T > t + l - 1$. At that point in time, the accumulated retirement benefit is converted into an annuity. The pension benefits are paid at the beginning of each year starting in year T, and the annual payment equals a fraction α of the average wage if one has participated for 40 years and scaled proportionally otherwise, accounted for the indexation policy of the pension fund. In the numerical applications, α has been set to 70%. Then the total payoff of l years participation equals

$$\sum_{\tau=T}^{\infty} I_{(\tau<S)}(x,\tau)\left(\alpha\frac{l}{40}\right)\left(\frac{1}{l}\sum_{i=0}^{l-1} w_{t+i}\mathcal{I}(t+i,\tau)\right), \tag{5.15}$$

where S is the year in which an individual dies, and is therefore a stopping time. The indicator function $I_{(\tau<S)}(x,\tau)$ in (5.15) equals 1 if the individual with age x survives year τ, and it is zero otherwise. The second term in (5.15), $\left(\alpha\frac{l}{40}\right)$, accounts for the number of years that an individual has participated. If $l = 40$, the individual receives a fraction α of the (possibly indexed) average wage. The last term in (5.15) determines the average wage, accounting for the indexation granted by the pension scheme. $\mathcal{I}(t+i,\tau)$ denotes the inflation indexation provided from time $t+i$ to time τ. If we denote the (commodity) price level at time t by Π_t, then $\mathcal{I}(t,\tau)$ can be defined as

$$\mathcal{I}(t,\tau) = \prod_{s=t+1}^{\tau}\left(1 + h\left(FR_s\right)\frac{\Pi_s}{\Pi_t}\right), \tag{5.16}$$

where $h\left(FR_s\right)$ represents the indexation policy of the pension fund, depending on the funding ratio at time s[23]. Examples are $h\left(FR\right) = 0$ for a nominal pension scheme and

[23]The funding ratio has been defined in this chapter as the ratio of assets to the nominal value of the

$h(FR) = 1$ for the real counterpart. On the other hand, in this chapter we consider the conditionally indexed scheme as well. This implies that full indexation is given if the funding ratio is sufficiently high (i.e. $FR > U$), but no indexation is granted if the funding ratio is too low (i.e. $FR < L$). In between the fraction of inflation indexation is determined proportionally. Formally,

$$h(FR) = \begin{cases} 0 & , FR < L \\ \frac{FR-L}{U-L} & , FR \in [L, U] \\ 1 & , FR > L. \end{cases} \tag{5.17}$$

Before determining the value of the payoff, it is important to realize that the total payoff is additive in the payoffs of a single year participation in year $t+i$, i.e.

$$\frac{\alpha}{40} \sum_{\tau=T}^{\infty} I_{(\tau<S)}(x,\tau) \left[w_{t+i} \mathcal{I}(t+i,\tau) \right]. \tag{5.18}$$

This property is natural within the average wage system, but is no longer valid in a final wage system, in which back service issues come into play. i.e. the decision to participate an additional year is dependent on the previous wages earned. In the average wage system, these considerations are irrelevant and therefore, we can focus in this chapter on a single year participation of an individual within a collective pension scheme.

In order to value the payoff in (5.18), we specify a pricing kernel that is consistent with a simple financial market[24]. The relevant economic factors are assumed to the real interest rate ($R_t^{R(1)}$), inflation (π_t), and stock returns in excess of the nominal short rate (r_t). The dynamics are captured by a VAR(1) - model, in which we assume that the process for inflation and the real interest rate move independently. Finally, we assume that inflation and the real interest rate are independent from excess stock returns. Formally,

$$R_{t+1}^{R(1)} = \mu_R + \phi_R \left(R_t^{R(1)} - \mu_R \right) + \varepsilon_{t+1}^R \tag{5.19}$$

$$\pi_{t+1} = \mu_\pi + \phi_\pi \left(\pi_t - \mu_\pi \right) + \varepsilon_{t+1}^\pi \tag{5.20}$$

$$r_{t+1} = \mu_r + \varepsilon_{t+1}^r, \tag{5.21}$$

liabilities.

[24]The model is similar, but not identical to the market specified in Campbell and Viceira (2001) or in Brennan and Xia (2002). They model the realized inflation as the sum of an expected and an unexpected inflation component, where the expected inflation is characterized by an AR(1) process. Instead, we model the realized inflation without decomposition, similar to Ang and Bekaert (2005). Ang and Bekaert (2005) modeled the inflation process as an ARMA(1,1), however, we did not find evidence for that. The yearly realized inflation was found to be best characterized by an AR(1) process based on Dutch yearly inflation data, provided by the CBS Netherlands.

with

$$\varepsilon_{t+1} \equiv \left(\varepsilon_{t+1}^{R}, \varepsilon_{t+1}^{\pi}, \varepsilon_{t+1}^{r}\right) \stackrel{i.i.d.}{\sim} N\left(0_{3\times 1}, diag\left(\sigma_R^2, \sigma_\pi^2, \sigma_r^2\right)\right), \tag{5.22}$$

and inflation is defined as $\pi_{t+1} \equiv \log \Pi_{t+1} - \log \Pi_t$. For the specification of the real pricing kernel (M_{t+1}), we postulate

$$-\log M_{t+1} = \alpha + \delta R_t^{R(1)} + \beta_R \varepsilon_{t+1}^R + \beta_r \varepsilon_{t+1}^r + \eta_{t+1}, \tag{5.23}$$

where $\eta_{t+1} \stackrel{i.i.d.}{\sim} N\left(0, \sigma_\eta^2\right)$ and η_s and ε_t mutually independent for all t and s. We refer to Campbell et al. (1997) for a motivation of such a pricing kernel. Note that an assumption underlying the kernel specification in (5.23) is that the 1-period inflation risk is not priced in real terms, in line with Ang and Bekaert (2005) and Campbell and Viceira (2001).[25] Stated differently, ε_{t+1}^{π} does not appear in the pricing kernel. In order to value nominal liabilities, we make use of the link between the nominal and the real pricing kernel

$$m_{t+1}^{\$} = m_{t+1} - \pi_{t+1}. \tag{5.24}$$

It is well-known that the affine nature of these models translates in affine nominal and real yields at all maturities and therefore, the corresponding bond prices are exponentially affine in the real rate and inflation, see for instance Campbell and Viceira (2001). Formally, we obtain

$$P_t^{(n)} = \exp\left(-A_n - B_{n,1} R_t^{R(1)} - B_{n,2}\pi_t\right), \tag{5.25}$$

for the price of a nominal bond at time t with time to maturity n and

$$P_t^{R(n)} = \exp\left(-A_n^R - B_{n,1}^R R_t^{R(1)}\right), \tag{5.26}$$

for the price of a real bond at time t with time to maturity n.

Using the nominal pricing kernel, the price of any nominal payoff X_{t+1} at time t can be obtained via

$$P_t = \mathbb{E}_t\left(M_{t+1} X_{t+1}\right). \tag{5.27}$$

In the same spirit, the value of a single year participation within a collective pension scheme can be determined as

$$\frac{\alpha}{40}\mathbb{E}_t\left(\sum_{\tau=T}^{\infty} I_{(\tau<S)}(x,\tau) M_{t,\tau}^{\$}\left[w_t \mathcal{I}(t,\tau)\right]\right), \tag{5.28}$$

[25] There is an empirical evidence from the UK indexed guilt market that investors demand risk-premium for holding inflation-linked assets (see, for instance Evans, 1998), however, for simplicity we assume that inflation risk is not priced, because we do not observe inflation-linked instruments in the Netherlands.

with

$$M^{\$}_{t,T} = \prod_{s=t+1}^{T} M^{\$}_{s}. \qquad (5.29)$$

Important elements in calculating this expectation are the dependencies between S, the time at which the individual dies, the financial markets, and the pricing kernel. In doing this, we make the common assumption that S and the financial market are independent. Secondly, we assume that the time of death, S, is independent of the nominal pricing kernel. This assumption is somewhat more subtle. If we consider a large collective pension scheme, then idiosyncratic risks in the individual life times will be negligible as an application of the law of large numbers. However, when survival probabilities of the participants as a whole increase due to improvements in medical care, this does constitute an important risk factor for the pension fund. Since we are not able to identify the 'price of mortality risk', we assume throughout that mortality risk is not priced, implying that the conditional expectation for in (5.28) factorizes into

$$\frac{\alpha}{40} \sum_{\tau=T}^{\infty} \mathbb{E}_t \left(I_{(\tau<S)}(x,\tau) \right) \cdot \mathbb{E}_t \left(M^{\$}_{t,\tau} \left[w_{t+i} \mathcal{I}(t+i,\tau) \right] \right). \qquad (5.30)$$

Appendix 5.B discusses in detail how we model $\mathbb{E}_t \left(I_{(\tau<S)}(x,\tau) \right)$. The second part can be valued using the specification of the financial market as presented before, in conjunction with the pricing kernel. When the pension schemes are straight nominal or real, the second conditional expectation can be determined easily using the closed-form solutions that result from the affine term structure model. In case of conditionally indexed pension schemes, we use Monte Carlo techniques to determine this value. In this simulation procedure, standard variance reduction methods, like control variate and antithetic variables, turn out to be useful to reduce the Monte Carlo error.

We calibrated the parameters of the financial market in such a way that reflects the main stylized facts in the observed nominal yield, inflation and stock market return data in the Netherlands. The 1-year nominal yield between 1975 and 2004 is proxied by the the Dutch 1-year euro (previously guilder) interest rate swap middle rate[26] downloaded from Datasteam, and the yearly inflation rate between 1975 and 2004 was supplied by the CBS

[26]The zero-coupon yield data are available for the period starting only from year 1997, which is very short to estimate its time-series properties. The euro/guilder interest rate swap market might contain some counterparty risk, however, the depth and the quality of the market in London is likely to make the counterparty risk limited. The comparison of the zero-yield with the swap rate in the period between 1997 and 2004 yielded a deviation of at most 0.1% point, also suggesting, that the swap rate is likely to be a good proxy.

Netherlands. The observed nominal term structure has the following characteristics. The first order autocorrelation of the nominal 1-year rate is about 0.8 if the yearly inflation is also included in the regression as an explanatory variable, which has a coefficient of -0.08. The autocorrelation coefficient of the yearly inflation was estimated to be in the magnitude of 0.75. The standard deviation of the nominal annual 1-year interest rates and inflation rates are 1.8% and 1.1% respectively, with a correlation coefficient of about 0.6. The mean of the 1-year nominal yield is in the order of magnitude of 5.6%, and for the yearly inflation the corresponding value is about 2.3%. Moreover, we estimated a 1.2% term premium on a nominal bond with a maturity of 50 years by using a single factor affine Gaussian term structure model driven by the 1-period nominal rate[27].

In order to match the above mentioned characteristics of the observed nominal term structure to a large extent, the autocorrelation coefficient of the real 1-year yield was chosen to be 0.85 with a mean of 3.3% and a standard deviation of 1.6%.

The excess stock return is about 6%[28] with a standard deviation of 24% p.a., based on the total return index of the Dutch market downloaded from Datastream for the period between 1983 and 2004.

Because of the long-term nature of the pension claims, the correct representation of the long end of the term structure is far more important than that of the short end. The model-implied long rates at the beginning of 2004 were below the observed long rates at that time. To fix this problem the factors have been rotated where the nominal 10-year rate takes the role of the 1-year rate and the observed 10-year rate is taken as input for the analysis.

[27]The market price of risk parameter was calibrated by observing the 10-year nominal yield with error, and the 10-year rate between 1975 and 2004 was proxied by the 10-year benchmark yield provided by Datastream. For more methodological details, see Ang and Piazzesi (2003) for instance.

[28]Fama and French (2002) suggest that the equity premium estimated from fundamentals (for instance, the dividend or earnings growth rates) can be much lower than the equity premium produced by the average stock return. For simplicity, to calculate the excess return we used the average stock return in the sample from 1983 and 2004 and no fundamentals.

Chapter 6

Conclusions and Directions for Further Research

6.1 Summary and conclusions

After summarizing the literature in Chapter 2, in Chapter 3 we proposed an alternative mortality model which generalizes the empirically estimated version of the Lee and Carter model by allowing for a time-varying trend, depending on a few underlying factors. The proposed method aims to make the estimated future trend less sensitive to the sample period, which was illustrated. The model is written on a state space form and estimated with quasi maximum likelihood using the Kalman filtering method. Several model specifications were considered. We fitted the models for the Dutch male population. The model version with the 2-factor moving average latent process was considered to be the one with the best fit. Mortality forecasts were produced for selected groups with confidence intervals including both macro-longevity and parameter risk.

Chapter 4 quantifies the effect of mortality improvement and mortality risk on future solvency positions of pension funds. We first looked at the importance of mortality improvement on the expected remaining lifetime and the present values of annuities. Then we quantified the potential effect of longevity risk, including micro-, macro-longevity, and parameter risk on the solvency requirements of annuity funds for several future horizons by assuming no market risk. The results imply that a large fund portraying the Dutch population needs to have an initial funding ratio of 103.5% in a 1-year horizon and a 107.1% in a 5-year horizon in order to set the probability of underfunding to 2.5%, if we take into account all uncertainty in survival, and we assume no market risk. Moreover, the chapter presents an extension, where we allow for randomness in future realization of the term structure of interest rates and future stock returns. We analyze several asset compositions. If the interest rate sensitivity of the asset portfolio matches the interest rate sensitivity of the liabilities and no stock investments are considered, the longevity risk dominates the total uncertainty in the future distribution of the funding ratio. As soon as the interest rate hedge becomes less and less perfect, and an increasing fraction of the asset portfolio consists of stocks invested in the market portfolio, the relative size of longevity risk decreases.

Chapter 5 quantifies the incentives for individuals that participate in a collective pension scheme which disregards heterogeneity in survival characteristics of the policyholders in pricing. We show that young males with low education have a strong incentive to opt-out of the collective system in case of uniform pricing, since their contribution is high relative to the benefit obtained. This incentive is enforced by the fact that the switching costs for young individuals are relatively low. Moreover, it turns out that the indexation quality of the scheme is a non-negligible determinant of the incentives provided to participants.

6.2 Some directions for further research

The underlying latent factors which drive the mortality rates of the Dutch population were separately identified for men and women. However, our estimation results showed that there is a substantial correlation among those factors. This might reflect that a large part of the variation in the male and female mortality is explained by common factors. As an extension of the model presented in Chapter 3, the male and female mortality can be estimated simultaneously, driven by the same set of factors. Similarly, there is likely to be a comovement among the mortality of similar countries, such as in the EU. An estimation of a multifactor mortality model which drives the mortality of several countries might have a practical relevance for multinational companies in the insurance and pension industry.

The mortality model we considered has the potential of both future mortality improvement and mortality deterioration. We believe that we cannot exclude the risk in mortality deterioration in the future, which would significantly affect the risk of the portfolio of life insurance companies. However, given the downward sloping trend in the future, improvement is more likely than deterioration. The construction of a model which implies improvement with a large probability, and deterioration with a smaller probability is a topic for further research.

The fact that macro-longevity risk is not diversifiable raises the issue of a non-zero mortality risk premium in the price of survival related financial products. Throughout the thesis we assumed that markets are neutral to macro-longevity risk, implying a zero mortality risk premium, since we were not able to estimate it from observed data. If the payoff of survival related securities depends on traded assets and the market is complete, it can be perfectly replicated, the market price of mortality risk is uniquely identified. The literature has developed several approaches to price contingent claims in incomplete markets, which might be able to help to derive arbitrage-free prices of mortality linked financial securities. We could potentially use a method suggested by Carr et al. (2001), for instance. Carr et al. (2001) present a new pricing approach that bridges standard arbitrage pricing and expected utility maximization. Cochrane and Saá-Requejo (2000) derive bounds on asset prices when one cannot find a perfect replicating portfolio. However, the choice of the appropriate method developed by the incomplete market literature and its adaptation to our problem are fairly complex.

Chapter 5 analyzes the present value of participation for individuals in a collective scheme for one single year in the Netherlands. Only old-age pension has been considered and part of the money's worth differentials that have been identified can be compensated by the partner pension arrangements that are usually also included in actual pension products. Moreover we made the strong assumption that only the market value of what is received is relevant because agents can and will unwind all product features that are imposed by the pension fund. Subsequent analysis will have to consider the question to what extent these assumptions dominate the analysis.

Survival heterogeneity of the population is a key issue in identifying the direction and the size of the incentives of socioeconomic groups participating in a collective pension plan. In order to get a more precise and detailed picture, we need to estimate the heterogeneity by using recent Dutch data (so far we approximated it by the results of a survey conducted in Belgium). Moreover, the more subtle distinction of socioeconomic groups based on observable characteristics driving the differences in expected lifetime is important. For instance, apart from gender, age, and education, other characteristics, such as wealth, occupation, health status etc. are also important determinants of the expected lifetime of an individual.[1]

Insurers of the non-collective arrangements are faced with the potential of adverse selection, which makes the differences among the money's worth measures even more

[1]It is inevitable that some of these characteristics are interrelated, for instance, higher educated are likely to have higher wealth, better occupation, and better health status, but not necessarily.

substantial, since the self-selection of high-cost policyholders (from the insurer point of view) increases the overall contribution rate. It might lead to an even higher heterogeneity in the money's worth measures which induce stronger incentives of low-cost to opt out.

References

Ang, A. and G. Bekaert (2005), The term structure of real rates and expected inflation, Working Paper.

Ang, A., G. Bekaert, and J. Liu (2005), Why stocks may disappoint, *Journal of Financial Economics*, 76, 471–508.

Ang, A. and M. Piazzesi (2003), A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics*, 50, 745–787.

Auerbach, A. J., L. J. Kotlikoff, and W. Leibfritz (1999), *Generational accounting around the world*, NBER, Chicago University Press.

Bell, W. R. (1984), An introduction to forecasting with time series models, *Insurance: Mathematics and Economics*, 3, 241–255.

Benjamin, B. and J. H. Pollard (1993), *The analisys of mortality and other actuarial statistics*, Institute of Actuaries and Faculty of Actuaries.

Benjamin, B. and A. S. Soliman (1993), *Mortality on the move*, Institute of Actuaries, Oxford.

Biffis, E. (2005), Affine processes for dynamic mortality and actuarial valuations, *Insurance: Mathematics and Economics*, 37, 443–468.

Blake, D., A. J. G. Cairns, and K. Dowd (2006), Living with mortality: Longevity bonds and other mortality-linked securities, Working Paper, presented to the Faculty of Actuaries, 16 January 2006.

Blaschke, E. (1923), Sulle tavole di mortalita variabili col tempo, *Giornale di Mathematica Financziaria*, 5, 1–31.

Booth, H., J. Maindonald, and L. Smith (2002), Age-time interactions in mortality projection: Applying Lee-Carter to Australia, Working Papers in Demography, Research School of Social Sciences, The Australian National University.

Bos, V., A. E. Kunst, J. Garssen, and J. P. Mackenbach (2005), Socioeconomic inequalities in mortality within ethnic groups in the Netherlands, 1995-2000, *Journal of Epidemiology and Community Health*, 59, 329–335.

Brennan, M. J. and Y. Xia (2002), Dynamic asset allocation under inflation, *The Journal of Finance*, 57, 1201–1238.

Breusch, T. S. and A. R. Pagan (1980), The Lagrange multiplier test and its applications to model specification in econometrics, *The Review of Economic Studies*, 47, 239–253.

Brouhns, N., M. Denuit, and I. Van Keilegom (2005), Bootstrapping the Poisson log-bilinear model for mortality forecasting, *Scandinavian Actuarial Journal*, 2005, 212–224.

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

Brown, J. R. (2002), Differential mortality and the value of individual account retirement annuities, in M. Feldstein and J. B. Liebman, eds., *The distributional aspects of social security and social security reform*, pp. 401–440, The Chicago University Press.

Brown, J. R. (2003), Redistribution and insurance: mandatory annuitization with mortality heterogeneity, *The Journal of Risk and Insurance*, 70, 17–41.

Brown, J. R., J. B. Liebman, and J. Pollet (2002), Estimating life tables that reflect socioeconomic differences in mortality, in M. Feldstein and J. B. Liebman, eds., *The distributional aspects of social security and social security reform*, pp. 447–457, The Chicago University Press.

Cairns, A. J. G., D. Blake, and K. Dowd (2006a), Pricing death: Frameworks for the valuation and securitization of mortality risk, *ASTIN Bulletin*, 36, to appear.

Cairns, A. J. G., D. Blake, and K. Dowd (2006b), A two-factor model for stochastic mortality with parameter uncertainty, *Journal of Risk and Insurance*, in press.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira (2003), A multivariate model of strategic asset allocation, *Journal of Financial Economics*, 67, 41–80.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997), *The econometrics of the financial markets*, Princeton University Press.

Campbell, J. Y. and L. M. Viceira (2001), Who should buy long term bonds?, *The American Economic Review*, 91, 99–127.

Campbell, J. Y. and L. M. Viceira (2002), *Strategic asset allocation: Portfolio choice for long-term investors*, Oxford University Press.

Carone, G., D. Costello, N. D. Guardia, G. Mourre, B. Przywara, and A. Salomaki (2005), The economic impact of ageing populations in the EU25 Member States, Economic Papers 236, European Commission.

Carr, P., H. Geman, and D. B. Madan (2001), Pricing and hedging in incomplete markets, *Journal of Financial Economics*, 62, 131–167.

Carter, L. R. and A. Prskawetz (2001), Examining structural shifts in mortality using the Lee-Carter method, MPIDR Working Paper WP 2001-007.

Coale, A. and G. Guo (1989), Revised regional model life tables at very low levels of mortality, *Population Index*, 55, 613–643.

Coale, A. and E. E. Kisker (1990), Defects in data on old age mortality in the United States: new procedures for calculating approximately accurate mortality schedules and life tables at the highest ages, *Asian and Pacific Population Forum*, 4, 1–31.

Cocco, J. F. and P. Lopes (2004), Defined benefit or defined contribution? An empirical study of pension choices, Working Paper.

Cochrane, J. H. (2001), *Asset pricing*, Princeton University Press.

Cochrane, J. H. and J. Saá-Requejo (2000), Beyond arbitrage: Good-deal asset price bounds in incomplete markets, *Journal of Political Economy*, 108, 79–119.

Coppola, M., E. Di Lorenzo, and M. Sibillo (2000), Risk sources in a life annuity portfolio: Decomposition and measurement tools, *Journal of Actuarial Practice*, 8, 43–61.

Coppola, M., E. Di Lorenzo, and M. Sibillo (2003a), Further remarks on risk sources: Measuring the case of a life annuity portfolio, *Journal of Actuarial Practice*, 10.

Coppola, M., E. Di Lorenzo, and M. Sibillo (2003b), Stochastic analysis in life office management: applications to large annuity portfolios, *Applied Stochastic Models in Business and Industry*, 19, 31–42.

Cossette, H., A. Delwarde, M. Denuit, F. Guillot, and E. Marceau (2005), Pension plan valuation and dynamic mortality tables, Working Paper, presented at The Ninth International Congress on Insurance: Mathematics and Economics, hosted by the École d'actuariat of Université Laval, Québec, Canada on July 6-8, 2005.

Cui, J., F. de Jong, and E. Ponds (2005), Intergenerational transfers within funded pension schemes, Working Paper.

Dahl, M. (2004), Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts, *Insurance: Mathematics and Economics*, 35, 113–136.

Deboosere, P. and S. Gadeyne (2002), Levensverwachting en sterftekansen naar geslacht, gewest en onderwijsniveau in Belgie, 1991-1996, *Bevolking en Gezin*, 31, 47–74.

Di Lorenzo, E. and M. Sibillo (2002), Longevity risk: Measurement and application perspectives, Proceedings of the "2nd Conference in Actuarial Science and Finance" Samos, 2002.

Dowd, K., A. J. G. Cairns, and D. Blake (2006), Mortality-dependent financial risk measures, *Mathematics and Economics*, in press.

Durbin, J. and S. J. Koopman (2001), *Time series analysis by state space methods*, Oxford University Press, Oxford.

Economic Policy Committee and European Commission (2005), The 2005 EPC projections of age-related expenditure (2004-2050) for the EU25 member states: underlying assumptions and projection methodologies, European Economy Reports and Studies 4, European Commission.

Evans, M. D. D. (1998), Real rates, expected inflation, and inflation risk premia, *The Journal of Finance*, 53, 187–218.

Fama, E. F. and K. French (2002), The equity premium, *The Journal of Finance*, 57, 637–659.

Feldstein, M. and J. B. Liebman (2002), The distributional effects of an investment-based social security system, in M. Feldstein and J. B. Liebman, eds., *The distributional aspects of social security and social security reform*, pp. 263–322, The Chicago University Press.

Finkelstein, A. and J. Poterba (2002), Selection effects in the UK individual annuities market, *The Economic Journal*, 112, 28–50.

Finkelstein, A. and J. Poterba (2004), Adverse selection in insurance markets: Policyholder evidence from the U.K. annuity market, *Journal of Political Economy*, 112, 183–208.

FOPI (2004), White Paper on the Swiss solvency test, Technical report, Swiss Federal Office of Private Insurance (FOPI).

Forfar, D. O., M. A. McCutcheon, and A. D. Wilkie (1988), On graduation by mathematical formula, *Journal of the Institute of Actuaries*, 115, 1–149.

Gerber, H. U. (1997), *Life insurance mathematics*, Springer, Zürich.

Girosi, F. and G. King (2005a), Demographic Forecasting, Unpublished manuscript.

Girosi, F. and G. King (2005b), A reassessment of the Lee-Carter mortality forecasting method, Working Paper, Harvard University.

Gollier, C. (2005), Intergenerational risk-sharing and risk-taking of a pension fund, Presented at the Conférence en l'honneur de Jean-Jacques Laffont, Toulouse, France, June 30 - July 2, 2005.

Gompertz, B. (1825), On the nature of the law of human mortality and on a new method of determining the value of life contingencies, *Phil. Trans. of Royal Society*, 115, 513–585.

Hamilton, J. D. (1994), *Time series analysis*, Princeton University Press, Princeton, New Jersey.

Heligman, L. and J. H. Pollard (1980), The age pattern of mortality, *Journal of the Institute of Actuaries*, 107, 49–80.

Herten, v. L. M., K. Oudshoorn, R. J. M. Perenboom, Y. M. Mulder, N. Hoeymans, D. J. H. Deeg, F. R. M. Portrait, A. E. Kunst, F. van Lenthe, J. C. M. Witterman, C. T. M. van Rossum, and A. J. M. van Loon (2002), Gezonde levensverwachting naar sociaal economische status, TNO-rapport, 2002.170.

Huisman, M., A. E. Kunst, O. Andersen, M. Bopp, J.-K. Borgan, G. Costa, P. Deboosere, G. Desplanques, A. Donkin, S. Gadeyne, C. Minder, E. Regidor, T. Spadea, T. Valkonen, and J. P. Mackenbach (2004), Socioeconomic inequalities in mortality rates among elderly people in eleven European countries, *Journal of Epidemiology and Community Health*, 58, 468–475.

Huisman, M., A. E. Kunst, M. Bopp, J.-K. Borgan, C. Borrell, G. Costa, P. Deboosere, S. Gadeyne, M. Glickman, C. Marinacci, C. Minder, E. Regidor, T. Valkonen, and J. P. Mackenbach (2005), Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations, *The Lancet*, 365, 493–500.

Keyfitz, N. (1982), Choice of function for mortality analysis: Effective forecasting depends on minimum parameter representation, *Theoretical Population Biology*, 21, 329–352.

Khalaf-Allah, M., S. Haberman, and R. Verral (2006), Measuring the effect of mortality improvements on the cost of annuities, *Insurance: Mathematics and Economics*, in press.

Koissi, M. and A. F. Shapiro (2006), Fuzzy formulation of the Lee-Carter model for mortality forecasting, *Insurance: Mathematics and Economics*, in press.

Koissi, M., A. F. Shapiro, and G. Högnäs (2006), Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval, *Insurance: Mathematics and Economics*, 38, 1–20.

Kune, J. B. (2005), Billijkheid en doelmatigheid in het systeem van (aanvullende) pensioenvoorzieningen, Stichting PensioenWetenshappen (in Dutch).

Kunst, A. (1997), *Cross-national comparisons of socio-economic differences in mortality*, Ph.D. thesis, Erasmus University of Rotterdam.

Kwon, H. and B. L. Jones (2006), The impact of the determinants of mortality on life insurance and annuities, *Insurance: Mathematics and Economics*, 38, 271–288.

Lee, R. D. (2000), The Lee-Carter method for forecasting mortality, with various extensions and applications, *North American Actuarial Journal*, 4, 80–93.

Lee, R. D. and L. R. Carter (1992), Modeling and forecasting U.S. mortality, *Journal of the American Statistical Association*, 87, 659–671.

Lee, R. D. and T. Miller (2001), Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography*, 38, 537–49.

Lindbergson, M. (2001), Mortality among the elderly in Sweden 1988-97, *Scandinavian Actuarial Journal*, 1, 79–94.

Longstaff, F. A. and E. S. Schwartz (2001), Valuing American options by simulation: A simple least-squares approach, *The Review of Financial Studies*, 14, 113–147.

Mackenbach, J. P., V. Bos, O. Andersen, M. Cardano, G. Costa, S. Harding, A. Reid, O. Hemström, T. Valkonen, and A. E. Kunst (2003), Widening socioeconomic inequalities in mortality in six Western European countries, *International Journal of Epidemiology*, 32, 830–837.

Makeham, W. M. (1860), On the law of mortality, *Journal of the Institute of Actuaries*, 13, 325–358.

McNown, R. and A. Rogers (1989), Forecasting of mortality: A parameterized time series approach, *Demography*, 26, 645–660.

Milevsky, M. and S. D. Promislow (2001), Mortality derivatives and the option to annuitize, *Insurance: Mathematics and Economics*, 29, 299–318.

Mitchell, O., J. Poterba, M. Warshawsky, and J. Brown (1999), New evidence on money's worth of individual annuities, *American Economic Review*, 89, 1299–1318.

Nijman, T. E. and R. S. J. Koijen (2006), Valuation and risk management of inflation sensitive pension rights, in N. Kortleve, T. E. Nijman, and E. Ponds, eds., *Fair Value and Pension Fund Management*, Elsevier Publishers.

Olivieri, A. (2001), Uncertainty in mortality projections: an actuarial perspective, *Insurance: Mathematics and Economics*, 29, 231–245.

Olivieri, A. (2002), Risk and reinsurance in life annuity portfolios, Working Paper.

Olivieri, A. and E. Pitacco (2003), Solvency requirements for pension annuities, *Journal of Pension Economics and Finance*, 2, 127–154.

Pappas, G., S. Queen, W. Hadden, and G. Fisher (1993), The increasing disparity in mortality between socioeconomic groups in the United States: 1960 and 1986, *New England Journal of Medicine*, 329, 103–109.

Perks, W. (1932), On some experiments in the graduation of mortality statistics, *Journal of the Institute of Actuaries*, 63, 12–57.

Pitacco, E. (2002), Longevity risk in living benefits, Working Paper, Center for Research on Pension and Welfare Policies, Presented at the third annual CeRP conference "Developing an Annuity Market in Europe", Moncalieri, Turin, 21-22 June 2002.

Pitacco, E. (2004), Survival models in a dynamic context: a survey, *Insurance: Mathematics and Economics*, 35, 279–298.

Ponds, E. H. M. (2003), Pension funds and value-based generational accounting, *Journal of Pension Economics and Finance*, 2, 295–325.

Poterba, J. and D. Wise (1998), Individual financial decisions in retirement saving plans and the provision of resources for retirement, in M. Feldstein, ed., *Privatizing social security*, NBER, The Chicago University Press.

Preston, S. and P. Taubman (1994), Socioeconomic differences in adult mortality and health status, in L. G. Martin and S. H. Preston, eds., *Demography and aging*, pp. 279–318, National Academy Press, Washington DC.

Renshaw, A. E. and S. Haberman (2003a), Lee-Carter mortality forecasting with age-specific enhancement, *Insurance: Mathematics and Economics*, 33, 255–272.

Renshaw, A. E. and S. Haberman (2003b), On the forecasting of mortality reduction factors, *Insurance: Mathematics and Economics*, 32, 379–401.

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

Renshaw, A. E., S. Haberman, and P. Hatzopoulos (1996), The modelling of recent mortality trends in United Kingdom male assured lives, *British Actuarial Journal*, 2, 449–477.

Schrager, D. F. (2006), Affine stochastic mortality, *Insurance: Mathematics and Economics*, 38, 81–97.

Sithole, T. Z., S. Haberman, and R. J. Verrall (2000), An investigation into parametric models for mortality projections, with applications to immediate annuitants' and life office pensioners' data, *Insurance: Mathematics and Economics*, 27, 285–312.

Tabeau, E. (2001), A review of demographic forecasting models for mortality, in E. Tabeau, A. Van Den Berg Jeths, and C. Heathcote, eds., *Forecasting mortality in developed countries: Insights from a statistical, demographic and epidemiological perspective*, pp. 1–32, Kluwer Academic Publishers, Dordrecht.

Thiele, P. N. (1872), On a mathematical formula to express the rate of mortality throughout the whole of life, *Journal of the Institute of Actuaries*, 16, 313–329.

Wilmoth, J. R. (1993), Computational methods for fitting and extrapolating the Lee-Carter model of mortality change, Technical report, Department of Demography, University of California, Berkeley.

Samenvatting (Summary in Dutch)

In de loop van de 20^e eeuw is de levensverwachting in verschillende landen gestaag gestegen. Als dit fenomeen zich voortzet in de toekomst, dan zou de bevolkingssamenstelling drastisch kunnen wijzigen. De combinatie van langere levensduur en lagere fertiliteit impliceert een steeds groter wordend aandeel van ouderen. Deze *vergrijzing* beïnvloedt verschillende aspecten van de macroeconomie, zoals productiviteit, arbeidsaanbod, werkgelegenheid en economische groei. Deze thesis richt zich op de interactie tussen vergrijzing en de financiële markten.

Indien toekomstige sterftekansen met zekerheid bekend zouden zijn, dan zouden we ook met zekerheid de verwachte resterende levensduur kunnen bepalen. Dat neemt echter niet weg dat het aantal personen dat een bepaalde leeftijd zal bereiken nog steeds onzeker blijft. Het risico dat gepaard gaat met deze onzekerheid zullen we *micro-langlevenrisico* noemen. Het is duidelijk dat de wet van de grote aantallen ervoor zorgt dat indien we een voldoende grote groep individuen bekijken, micro-langlevenrisico verwaarloosbaar klein wordt. Echter, zoals blijkt uit historische gegevens, toekomstige sterftekansen zijn onzeker. Deze onzekerheid creëert een extra bron van risico die niet gereduceerd kan worden door een voldoende grote groep individuen samen te nemen. Dit risico zullen we in het vervolg *macro-langlevenrisico* noemen. Om dit risico te kunnen meten en beheersen, hebben we een model nodig voor het voorspellen van toekomstige sterftekansen. Dit vormt het centrale thema van de thesis. Vervolgens is het belangrijk om de onzekerheid met betrekking tot toekomstige sterfteontwikkeling correct te verwerken in premies voor annuïteiten. Pensioenfondsen en levensverzekeraars kunnen immers potentieel zware verliezen lijden indien zij de toekomstige sterfteontwikkeling foutief inschatten.

Hoofdstuk 1 omschrijft een aantal karakteristieken van geobserveerde sterfte-intensiteiten. Hoofdstuk 2 geeft een overzicht van de literatuur met betrekking tot het voorspellen van toekomstige sterftekansen, en het risico dat daaruit voortvloeit. Hoofdstuk 3 introduceert een model voor toekomstige sterftekansen. In het Lee en Carter model (Lee en Carter, 1992) worden de sterftekansen gemodelleerd als een tijdsinvariante leeftijdsspecifieke constante, plus het product van een tijdsafhankelijke latente factor en een leeftijdsafhankelijke tijdsinvariante factor. In hoofdstuk 3 nemen we de herformulering van Girosi en King (2005) als benchmark en generaliseren deze door het toelaten van een tijdsvariërende trend die afhankelijk kan zijn van een aantal factoren. Deze aanpassing van het Girosi en King-model heeft als doel de voorspellingen minder gevoelig te laten zijn voor de gekozen steekproefperiode. Het model wordt geschreven in de zogeheten toestandsruimtevorm, en wordt geschat met de quasi maximale aannemelijkheidsmethode en maakt gebruik van 'Kalman filtering'. Na het toetsen van verschillende specificaties bleek dat een 'two-factor moving average latent process' de beste fit genereerde. Dit model werd vervolgens gebruikt om voor verschillende leeftijdsgroepen betrouwbaarheidsintervallen voor de toekomstige sterftekansen te genereren, rekening houdend met parameterrisico.

Hoofdstuk 4 maakt gebruik van het in hoofdstuk 3 geïntroduceerde model om het belang van onzekere sterfteontwikkeling op de solvabiliteitspositie van een pensioenfonds te analyseren. Eerst wordt er gekeken naar het effect van toekomstige sterftekansontwikkeling op de verwachte resterende levensduur en de contante waarde van een annuïteit. Vervolgens kwantificeren we het effect van langlevenrisico op de solvabiliteitsvereiste voor een portefeuille van annuïteiten. In alle gevallen wordt rekening gehouden met micro-langlevenrisico, macro-langlevenrisico en parameterrisico. De resultaten tonen aan dat voor grote fondsen waarvan de leeftijdssamenstelling vergelijkbaar is met die van de gehele Nederlandse bevolking, de initiële dekkingsgraad in de orde van grootte van 103% (107%) moet zijn om de kans op onderdekking op een termijn van 1 jaar (5 jaar) te beperken tot maximaal 2,5%. Het langlevenrisico zou echter ook beperkt kunnen worden door het kopen van een zogenaamd stop-loss herverzekeringscontract. We gebruiken het model voor de voorspelling van sterftekansen voor het prijzen van zulke contracten. Ten slotte kijken we naar het effect van langlevenrisico op een portefeuille van annuïteiten indien er ook financieel risico aanwezig is. We analyseren hierbij verschillende beleggingssamenstellingen. Zodra de gevoeligheid voor rentewijzigingen van de bezittingen niet goed aansluit bij de gevoeligheid van de verplichtingen voor dergelijke wijzigingen zal het langlevenrisico relatief verwaarloosbaar zijn ten opzichte van het financiële risico.

Uit bestaande literatuur (bijvoorbeeld Kunst, 1997; Brown, 2002; Huisman et al., 2004) blijkt dat in verschillende Europese landen de levensduur van een individu afhankelijk is van geslacht en opleiding. Vrouwen leven *ceteris paribus* gemiddeld langer dan mannen en hoger opgeleiden leven *ceteris paribus* gemiddeld langer dan lager opgeleiden.

Indien een pensioengerechtigde een lagere verwachte levensduur heeft dan het gemiddelde in het fonds, en de pensioenpremie is gebaseerd op het gemiddelde, dan zal dit individu een te hoge prijs betalen voor het pensioen. In hoofdstuk 5 wordt de contante waarde berekend van een jaarbijdrage aan een collectief pensioenplan, afhankelijk van de specifieke karakteristieken van het individu. In vele landen hebben werknemers impliciet of expliciet de optie om niet deel te nemen aan het collectieve pensioenplan. Bijdragen zijn vaak onafhankelijk van leeftijd, geslacht en opleidingsniveau. Gezien de heterogeniteit in sterftekansen creëert dit prikkels om al dan niet deel te nemen. We tonen aan dat jonge mannen met lage opleiding een sterke prikkel hebben om niet deel te nemen, omdat hun bijdrage relatief hoog is ten opzichte van de waarde van hun pensioenuitbetalingen. Deze prikkel wordt nog versterkt door het feit dat de kosten van een overgang van het ene naar het andere pensioenplan voor jonge individuen relatief laag zijn. Bovendien blijkt dat de indexatiekwaliteit van het pensioenplan een niet verwaarloosbare factor is in de contante waarde van deelname.