Pension and Health Insurance: Phase-Type Modeling

Maria Govorun

Thèse présentée en vue de l'obtention
du grade de docteur en sciences

Août 2013

Université Libre de Bruxelles

Faculté des sciences

Maria Govorun
Université Libre de Bruxelles
Département d'Informatique
Blvd du Triomphe - CP 212
1050 Bruxelles
Belgique
Email: mgovorun@ulb.ac.be, govorun@mail.ru

Thèse de doctorat présentée en séance publique le 26 août 2013 à l'Université Libre de Bruxelles.

Jury: Gianluca Bontempi, Président du jury, ULB
Griselda Deelstra, Secrétaire du jury, ULB
Bernard Fortz, Promoteur, ULB
Guy Latouche, Co-promoteur, ULB
Xiaoming Liu, Western University, London, Canada
Stéphane Loisel, ISFA - Université Lyon 1, Lyon, France

Les Presses Universitaires de Bruxelles (P.U.B.), 2013.

ULB

Contents

Acknowledgments

I would like to express my deepest appreciation to all those who provided me the possibility to complete this PhD thesis. I offer my sincerest gratitude to my supervisor Guy Latouche for his ongoing support. Thank you, Guy, for the time that you always found for me in your busy schedule, thank you for your simple or complex, but always helpful advices, for our interesting discussions and for the moral support.

A big thanks goes to Stéphane Loisel with whom I had the opportunity to work during my studies. Thank you, Stéphane, for welcoming me at ISFA, for your practical advices and for your help in establishing connections with the scientific community.

I am thankful to Bruce Jones, Xiaoming Liu and David Stanford for the work experience I got during my stay at the Western University. I thank Xiaoming in particular for her comments that helped me to improve the thesis and gave me a clearer overview of my work. I would like to thank Sheldon Lin for inviting me to present my research at the University of Toronto.

I am heartily thankful to Griselda Deelstra for her time and her interest in my work. Thank you, Griselda, for opening to me interesting ideas of future research directions. I would also like to thank Marie-Ange Remiche for her help to obtain the grant for my studies and for her support.

I am grateful to the Ministère de la Communauté française de Belgique for the ARC grant AUWB-08/13–ULB 5. The grant offered me a good financial support during all years of my studies, including the coverage of numerous scientific trips.

I thank my colleagues from the Computer Science and Mathematics Departments for their sympathy and the good atmosphere I enjoyed at work. I thank my dear friends for their encouragements and help during the four years of these PhD times. I really appreciate all the great moments we had together.

Last, but not least, I would like to say a big thank you to all my family members as without them this PhD thesis would not have been possible. I am grateful to my family for supporting the idea of writing a PhD thesis in Belgium. I thank my dear mother for her patience and her support in the moments of doubts and frustrations. I am heartily thankful to my grandparents, babulya and dedushka, for their help in developing the character of never giving up. I thank my dear uncle and his family for always being there for me, simply.

Brussels, August 2013.

Foreword

For many years scientists of various research domains have been actively using phase-type methods to describe objects in the nature that can be characterized by different states. The best known applications of phase-type methods belong to telecommunication systems (in queueing theory, Neuts [45]), economics (Norberg [46]), health economics (Gardiner et al. [25]) and actuarial science (Bowers [10], Wolthuis [64]). It is an interesting fact that phase-type methods have also been used in baseball analysis (B. Bukiet and E. R. Harold [14]), algorithmic music composition (Franz [24]) and in social sciences (Acemoglu et al.[1]).

One of the main reasons of such wide applicability is given by the properties of phase-type distributions. The set of phase-type distributions is dense in the field of all positive-valued distributions, and therefore can be used to approximate any positive-valued distribution.

In 2007 Lin and Liu [40] argued for the use of phase-type distributions to model the life and health of an individual. Specifically, the authors define a finite-state continuous-time Markov process to represent the hypothetical aging process of an individual, this is called a *phase-type aging model* ("PH-aging model" in the sequel). Aging is described as a process of consecutive transitions from one health state to another until death. One important property of this model is that the states have some physical interpretation, and the number of states is not chosen arbitrarily, but is defined from data using a well specified algorithmic procedure. Another important property of the model, which makes it different from other phase-type models for health, and very relevant for actuarial applications at the same time, is that it provides a connection between the health state of an individual and his/her age.

Because of its nice properties, the model of Lin and Liu motivated us to develop phase-type methods in that part of actuarial science, where the life and the health of an individual play a central role. Clearly, it includes problems of life, pension and health insurance companies. The profitability of such companies naturally depends on the life duration and the health of insured individuals, which implies that these two characteristics of individuals have to be properly modeled.

The need to have a good model for an individual's health is reinforced by recent changes in human mortality. In most countries average life expectancy is steadily increasing as human mortality rates keep decreasing (Oeppen and Vaupel [47]). As emphasized in [47], the reduction of human mortality is difficult to predict: it is not the same for all ages, and it fluctuates in time. Apart from the increasing lifetime, one obvious outcome of the unpredicted mortality is that it changes the health development process of a human being. Clearly, this is a major concern for any life-linked insurance, because these changes may result in huge financial losses. The risk of losses due to unpredictably decreasing mortality is called *longevity risk*.

The thesis is organized into four parts. Part I is devoted to the concept of the phase-type lifetime. We introduce the necessary actuarial and phase-type background, and present the model of Lin and Liu [40] together with its stochastic analogue elaborated in Lin and Liu [41]. We perform the quantitative analysis of the states in the model and verify the applicability of the model to basic actuarial problems. We also investigate its applicability to estimate the longevity risk and to model correlated cohorts.

Whereas Part I may be viewed as introductory, Part II and Part III contain the main results of the thesis.

In Part II we employ the phase-type concept for life and health to examine the profitability of a pension fund. We first construct a profit-test model, where we focus on the *pre-retirement period* assuming that pension fund participants receive a lifetime annuity as a lump sum at the moment of retirement and disappear from our consideration. With this profit-test model, we examine the risks exposure of the pension fund. Specifically, we evaluate the risks related to the change in population dynamics and to the behavior of the financial market. We also examine the long-term profitability of the fund. We find it natural to assume that the health of pension fund participants after retirement significantly affects the profitability of the fund. This is our motivation to separately consider the *post-retirement period*, which is the focus of our attention at the end of Part II. Some of these results are presented in Govorun and Latouche [28], Govorun et al. [29] and Govorun et al. [30].

Part III is devoted to health insurance. Specifically, we are interested in the distribution of the *net present value* (abbreviated as "NPV") of health care costs for an individual of a given age. We examine the costs in both long and short term perspective that depends on the remaining lifetime of the individual. We propose and analyze both discrete and continuous time models for NPV allowing the health care cost to be dependent or independent of the health state. We study the effect of special events by performing different stress tests. For example, tests with respect to mortality rates allow us to study the financial impact of an increased lifetime spent in bad health states for which medical treatments are the most expensive.

In the end of the thesis we give our concluding remarks and briefly describe the directions of future research that we have thought of, in parallel to the main topics of this thesis.

We provide a more detailed introduction to each part in its own summary, as well as a detailed description of its structure and main results. Some useful algebraic properties are included in Appendix A.

The following notations are used throughout the text:

- Matrices are denoted by capital letters, I stands for the identity matrix;
- Row vectors are usually denoted by underlined letters, $\mathbf{0}$ and $\mathbf{1}$ stand for column vectors of zeros and ones, respectively;
- Superscript $^{\mathrm{T}}$ indicates the transposition operation;
- Sign $\stackrel{d}{=}$ stands for "equal in distribution";
- If M is a matrix,

  $M_{(i,\cdot)}$ is the ith line of M;

  $M_{(\cdot,i)}$ is the ith column of M.

Part I

Phase-type lifetime

Summary of Part I

We begin the thesis by introducing the assumption that the lifetime and health of an individual can be described by a phase-type distribution. We devote the first part of the thesis to the introduction and the motivation of this assumption because we employ it in Parts II and III to develop various mathematical models and obtain our main results.

In Chapter 1, we provide the background helpful in understanding the concept of the phase-type lifetime. It consists of actuarial definitions, the definitions of phase-type objects and their applicability in insurance. We also give a brief survey of known mortality models and recent mortality trends.

In Chapter 2 we present the PH-aging model of Lin and Liu [40], where the authors introduce the assumption of the phase-type lifetime. Specifically, for the construction of their PH-aging model the authors show how to specify a concrete phase-type distribution for the lifetime of an individual, they give some interpretation to the states and provide a parametrization procedure. We start our investigations by verifying that the PH-aging model can be used to solve complex actuarial problems. To do this, we show that the model gives satisfactory solutions to basic actuarial problems, such as pricing and modeling of the number of survivors in a population, in comparison with known classical methods.

The fact that the states in the PH-aging model, interpreted as health states, are not observable creates some difficulties in the tractability of the model. For this reason, we develop algorithms to better characterize the health states on the basis of available information about an individual.

We examine how the PH-aging model can be adapted to deal with the problem of an unpredicted increase in lifetime (termed "longevity" in the sequel). We ask a question "why do people live longer?" and show that the model allows us to distinguish two different causes of longevity: one is related to individual characteristics, the other is related to external environmental factors.

In view of the longevity problem, Lin and Liu in [41] introduce a stochastic analogue of the PH-aging model, which we present at the end of Chapter 2 and apply in Chapter 3.

Specifically, in Chapter 3, we investigate the theoretical applicability of the stochastic PH-aging model to the problem of correlated cohorts. The problem has been addressed in numerous studies. For example, in Cairns et al. [16] the authors argue the importance of modeling correlated cohorts for future mortality forecasting. They develop a two-population mortality model, which helps in making financial decisions related to longevity risk. Cairns et al. [15] propose a model to price and to design survival index-based financial instruments, which are called *longevity bonds*. The population models, considered in [15], [16] and many other papers, are parametric factor models, based on the modeling of correlated death rates for cohorts.

We suggest an approach to model the number of survivors in different cohorts as correlated random variables. Here, we combine the classical technique (Pollard [51]) to model the number of survivors as a Binomial random variable and the technique to correlate the cohorts via the survival rates. The survival rates are given by the stochastic PH-aging model, and we correlate them by assuming that all cohorts are subject to the same uncertainty in future survival rates, but initially they may have a different health distribution. Therefore, we have two sources of randomness: one is due to the Binomial assumption and one is due to the uncertainty in future survival rates. To investigate their impact on the number of survivors, we develop two other approaches, where we eliminate the sources turn by turn. In the end of the Chapter, for comparison purposes, we suggest another approach that assumes that all cohorts are subject to the same common shock.

As an illustration where it is useful to model the number of survivors, we present the problem of hedging a fixed cohort of pensioners with longevity bonds, which was considered in Leppisaari [38]. We give a numerical example to demonstrate how our approaches can be applied to this problem.

We indicate several main contributions of Part I. As our first contribution, we have validated the PH-aging model for pricing and modeling the number of survivors in a cohort.

Our second contribution is the development of a procedure, which allows one to translate information pertaining to calendar ages into information pertaining to the health states in the PH-aging model. This has lead to some practical and theoretical results. For example, if a cost value is available for each age, our procedure determines a cost value for each health state. Furthermore, If a cost value is known for an individual, we reduce the uncertainty in the distribution of health states for this individual. Thus, this procedure allows us to better characterize the health state of the individual.

As other contributions we indicate the development of various models aimed to estimate the longevity risk and to represent correlated cohorts, assuming that the lifetime and health are described by the PH-aging model or its stochastic analogue.

Chapter 1

Background

This chapter provides the necessary background to the assumption of the phase-type lifetime and to its applications in pension and health insurance. Sections 1.1 and 1.2 are devoted to actuarial and phase-type definitions, Section 1.3 is devoted to mortality modeling. In Section 1.4 we present some applications of phase-type methods in insurance that are relevant for our work.

1.1 Actuarial definitions

Actuarial science is the discipline that applies mathematical and statistical methods to assess risks in insurance and financial industries (see [63]). A very basic object in actuarial science is a *Mortality Table* . A mortality table may be defined as a set of mortality rates q_x for all the ages x, the conditional probabilities to die within one year given that one has survived until age x. Such tables also give other basic quantities l_x and d_x, where l_x is equal to the expected number of survivors to age x from the l_0 newborns and d_x is the expected number of deaths over each age interval $(x, x+1]$.

We use the standard notation: ${}_tq_x$ is the conditional probability to die before age $x+t$ and ${}_tp_x$ is the conditional probability to survive to the age $x+t$ given the individual has survived to the age x. Obviously, they satisfy the equations that

$$ {}_tp_x = p_x p_{x+1} \cdots p_{x+t-1}, \qquad {}_tq_x = 1 - {}_tp_x. \tag{1.1} $$

Here we use the conventional notation ${}_1p_x = p_x$. We illustrate the survival probability ${}_tp_x$ with respect to t for different x in Fig. 1.1. The data is taken from the mortality table 1911 for Swedish male population (referred in the sequel as SW1911M). We use this mortality table in all our numerical examples, unless stated otherwise, as Swedish mortality database provides the most detailed mortality and population data and is widely used by researches, for example, Lin and Liu [40]. In Fig. 1.1 we show ${}_tp_x$ for the ages from 10 to 70 with a 5 years step.

Another important parameter related to the mortality environment is the instantaneous rate of mortality or the *force of mortality* , μ_x, which is given by

$$ \mu_x = \lim_{h \to 0+} \frac{{}_hq_x}{h}. \tag{1.2} $$

It is shown (Bowers [10], Scott [54]) that

$$ {}_tp_x = \exp\left(- \int_0^t \mu_{x+r} dr \right). $$

The *central death rate* over the interval from $(x, x+1]$ is given by

$$ m_x = \frac{l_x - l_{x+1}}{L_x}, \tag{1.3} $$

where $L_x = \int_0^1 l_{x+t} dt$ is interpreted as the total expected number of years lived between ages x and $x+1$ by survivors from the l_0 newborns.

The *discount factor* v comes in two versions. According to Gerber [26], for a fixed annually compounded *interest rate*, δ, the discount factor is given by

$$ v = \frac{1}{(1+\delta)}. \tag{1.4} $$

For a fixed continuously compounded interest, we have

$$ v = e^{-\delta}, \tag{1.5} $$

in this case, δ is called the *force of interest.*

Consider a simple *life insurance* contract, which pays a sum equal to 1 at the end of the year of death of an individual aged x. Denote by Z the present value of the

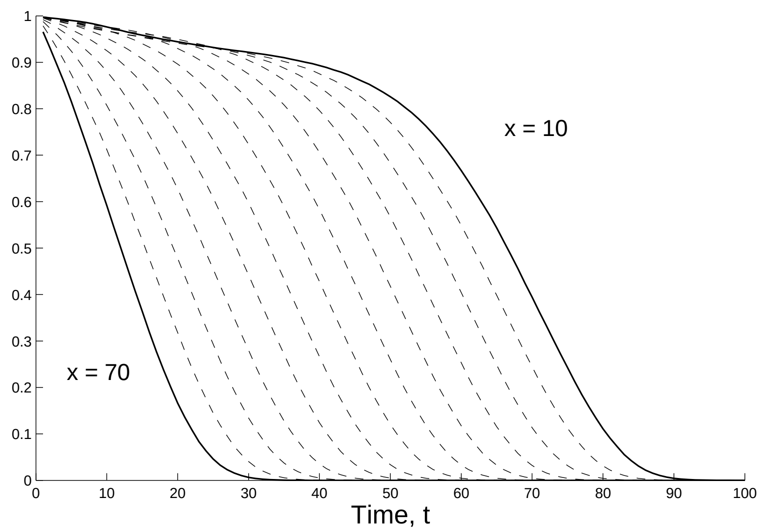

Figure 1.1: Survival probability ${}_tp_x$, SW1911M

insured sum. The expectation of Z is called *single benefit premium* (Gerber [26]). For this contract, $Z = v^{K+1}$, where K is the number of future full years of life of the individual. It is clear that

$$\mathrm{P}[\, Z = v^{k+1} \,] = \mathrm{P}[\, K = k \,] = {}_kp_x\, q_{x+k}, \quad k = 0, 1, 2, ... \tag{1.6}$$

Thus, the single benefit premium of a life insurance contract is given by

$$A_x = E[Z] = \sum_{k=0}^{\infty} v^{k+1}{}_kp_x\, q_{x+k}. \tag{1.7}$$

Consider a *term insurance* contract for n years which pays a unit sum upon the death of an individual of age x if it occurs before the term of the policy. Assuming that payment of the insured sum is made at the end of the year of death, we have

$$Z = \begin{cases} v^{K+1}, & K = 0, 1, ..., n-1 \\ 0, & K = n, n+1, ... \end{cases} \tag{1.8}$$

and the single benefit premium is denoted as

$$A^1_{x:\overline{n}|} = E[Z] = \sum_{k=0}^{n-1} v^{k+1}{}_kp_x\, q_{x+k} \tag{1.9}$$

A term insurance contract which pays a sum insured of 1 upon the survival of an individual of age x at the term of the policy is called *pure endowment*. For this type of insurance

$$Z = \begin{cases} 0, & K = 0, 1, ..., n-1 \\ v^n, & K = n, n+1, ... \end{cases} \tag{1.10}$$

The single benefit premium is defined by

$$A_{x:\overset{\,1}{\overline{n}|}} = E[Z] = v^n\, {}_np_x \tag{1.11}$$

Lifetime annuity pays the amount of 1 every year as long as the individual is alive. The present value of these payments is a random variable given by

$$\ddot{a}_{\overline{K+1}|} = 1 + v + v^2 + ... + v^K.$$

Mathematically, the value of lifetime annuity for an individual aged x is defined by

$$\ddot{a}_x = E[\ddot{a}_{\overline{K+1}|}] = \sum_{k=0}^{\infty} v^k\, {}_kp_x. \tag{1.12}$$

Similarly, *term annuity* is defined by

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^k\, {}_kp_x. \tag{1.13}$$

The *premium* or tariff is the price of an insurance contract. Premiums that do not include expenses related to the contract are called *net premiums*, whilst premiums which are actually charged to the client are called *gross premiums* (Scott [54]). Premiums are usually calculated from the equivalence principle, which states that the expectation of total loss of the insurer should be equal to zero. The net annual premiums computed for different types of insurances for an individual aged x at the start of the contract are presented below:

	Life insurance	*Term insurance*	*Pure endowment*
Net premium:	$A_x/\ddot{a}_x$	$A^1_{x:\overline{n}|}/\ddot{a}_{x:\overline{n}|}$	$A_{x:\overset{\,1}{\overline{n}|}}/\ddot{a}_{x:\overline{n}|}$

The annual gross premium for term insurance is defined by

$$(A^1_{x:\overline{n}|} + c_1)/((1-c_2)\,\ddot{a}_{x:n}),$$

where c_1 and c_2 are the initial and annual expenses per policy, respectively.

Assume that the mortality is not the single decrement of the model. Specifically, we assume that there are 2 independent decrements, X and Y. In accordance with Scott [54], let us make the following notations:

${}_t(ap)_x$	is the probability that an individual aged x will survive for at least t years with respect to X and Y
${}_t(aq)_x$	is the probability that an individual aged x leaves the system within t years due to either X or Y

As one assumes that the survival times with respect to the decrements X and Y are independent, it follows that

$$ {}_t(ap)_x = {}_tp_x^X \, {}_tp_x^Y , \qquad {}_t(aq)_x = 1 - {}_t(ap)_x, \tag{1.14} $$

where ${}_tp_x^X$ is the single-decrement probability for an individual aged x to survive with respect to the cause X only, ${}_tp_x^Y$ – with respect to the cause Y only. Therefore, for the term insurance, for example, the single benefit premium and the term annuity take form

$$ A^1_{x:\overline{n}|} = \sum_{t=0}^{n-1} v^{t+1} \left({}_t(ap)_x - {}_{t+1}(ap)_x \right) , \qquad \ddot{a}_{x:\overline{n}|} = \sum_{t=0}^{n-1} v^t \, {}_t(ap)_x . \tag{1.15} $$

1.2 Phase-type objects

We give the definition of a Markov chain introduced in Ross [52]. Consider a stochastic process $\{X_n, \ n = 0, 1, 2, \ldots\}$ that takes a finite or a countable number of possible values. If $X_n = i$, then the process is said to be in state i at time n. We suppose that, whenever the process is in state i, there is a fixed probability p_{ij} that it will next be in state j independently of its past. That is, we suppose that

$$ \mathrm{P}\left[\, X_{n+1} = j \mid X_n = i, \ X_{n-1} = i_{n-1}, \ldots, \ X_0 = i_0 \, \right] = p_{ij}, \tag{1.16} $$

for all states $i_0, \ i_1, \ldots, \ i_{n-1}, \ i, \ j$ and all $n \geq 0$. Such stochastic process is called a *Markov chain*. Let us denote the state space of this Markov chain by E. Since probabilities are nonnegative and since the process must take a transition into some state, we have that

$$ p_{ij} \geq 0, \quad \sum_{j \in \mathrm{E}} p_{ij} = 1, \quad i, j \in \mathrm{E}. \tag{1.17} $$

A probability distribution $\{\pi_j,\ j \in \mathrm{E}\}$ is said to be a *stationary distribution* for the Markov chain if

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij},\ j \geq 0.$$

The state j is *positive recurrent* if the expected number of transitions needed to return to this state is finite. If the stationary distribution exists and there is no other stationary distribution then it can be found from the equation

$$\pi_j = \lim_{n\to\infty} (P^n)_{ij}\,, \tag{1.18}$$

for all j, where P is the one step *transition probability matrix* $P = \{p_{ij}\}$, for all $i,\ j$. The stationary probability vector $\underline{\pi}$ can be found from the system

$$\begin{cases} \underline{\pi}P = \underline{\pi}, \\ \underline{\pi}\mathbf{1} = 1. \end{cases} \tag{1.19}$$

Denote the phase distribution vector at time t by $\underline{p}_t$. The dynamic of $\underline{p}_t$ is given by

$$\underline{p}_t = \underline{p}_0 P^t, \quad t \in \mathbb{N} \tag{1.20}$$

where $\underline{p}_0$ is a row vector representing the *initial distribution*.

In continuous time, the stochastic process $\{X_t\ :\ t \in \mathbb{R}^+\}$ is called a *Markov process* if and only if it satisfies the Markov property

$$\mathrm{P}[\,X_{t+s} = j \,|\, X_u, 0 \leq u \leq t\,] = \mathrm{P}[\,X_{t+s} = j \,|\, X_t\,], \tag{1.21}$$

for all $s, t \in \mathbb{R}^+$ and all j. The Markov process is characterized by matrix $P(t)$, which represents transition probabilities over an interval of time of length t. Let us also denote the *infinitesimal transition generator* of the system by Π. One method to find $P(t)$ is to solve the *Kolmogorov backward and forward equations*, which are defined by

$$\frac{d}{dt}P(t) = \Pi P(t), \qquad \frac{d}{dt}P(t) = P(t)\Pi. \tag{1.22}$$

According to Theorem 4.13 of Çinlar [18], the solution is given by

$$P(t) = e^{t\Pi}, \quad t \geq 0,$$

where $e^{t\Pi}$ is the *matrix exponential*, which is defined by

$$e^{t\Pi} = \sum_{n=0}^{\infty} \frac{1}{n!}(t\Pi)^n. \tag{1.23}$$

Therefore, the dynamic of $\underline{p}_t$ in continuous time is given by

$$\underline{p}_t = \underline{p}_0 e^{\Pi t}, \quad t \in \mathbb{R}^+. \tag{1.24}$$

The Markov process $\{X_t : t \in \mathbb{R}^+\}$ on the state space $\mathcal{E}$ is *irreducible* if, for all $i, j \in \mathcal{E}$, there is a path that leads from state i to state j and from state j to state i in the transition graph.

Phase-type distributions are described in details in Latouche and Ramaswami [36]. Here, we give the definition of a continuous-time phase-type distribution to the extent that it is relevant for this thesis. Consider a Markov process on the states $\{0, 1, \ldots, n\}$ with initial probability vector $(\alpha_0, \underline{\alpha})$ and infinitesimal generator Q given by

$$Q = \begin{bmatrix} \Upsilon & \underline{t}^{\mathrm{T}} \\ \underline{0} & 0 \end{bmatrix},$$

where $\underline{\alpha}$, $\underline{t}$, $\underline{0}$ are row vectors of size n and Υ is an $n \times n$ matrix. Since Q is the generator of a Markov process, we have that

$$\Upsilon_{ii} < 0, \qquad t_i \geq 0, \qquad \Upsilon_{ij} \geq 0 \qquad \text{for } 1 \leq i \neq j \leq n$$

and

$$\Upsilon \mathbf{1} + \underline{t}^{\mathrm{T}} = \mathbf{0}. \tag{1.25}$$

Also, for the initial probability vector we have

$$\alpha_0 + \underline{\alpha}\mathbf{1} = 1.$$

Then, the distribution of the time till absorption into state 0 is called the *phase-type distribution* with representation $(\underline{\alpha}, \Upsilon)$ and is denoted as $PH(\underline{\alpha}, \Upsilon)$.

Phase-type distributions have nice mathematical properties that allow to express different practical quantities of interest. Let us consider a random variable representing the lifetime of an individual, which we denote by L. As phase-type distributions are dense, we can assume L to have the representation $PH(\underline{\alpha}, \Upsilon)$. Then, the *survival function* is defined by

$$S(x) = \underline{\alpha} e^{\Upsilon x} \mathbf{1}, \quad x > 0, \tag{1.26}$$

where $e^{\Upsilon x}$ is the matrix exponential defined by Eq. (1.23). The corresponding *phase-type probability density function* is given by

$$f(x) = \underline{\alpha} e^{\Upsilon x} \underline{t}^{\mathsf{T}}, \quad x > 0. \tag{1.27}$$

In continuous time, the expected lifetime is determined by

$$E[L] = -\underline{\alpha} \Upsilon^{-1} \mathbf{1}. \tag{1.28}$$

In discrete time, it is determined by

$$E[L] = \underline{\alpha}(I - e^{\Upsilon})^{-1}\mathbf{1}, \tag{1.29}$$

and the *generating function* is

$$g(x) = x\underline{\alpha}(I - xe^{\Upsilon})^{-1}(\mathbf{1} - e^{\Upsilon}\mathbf{1}). \tag{1.30}$$

Another phase-type object that we use in this thesis is a *fluid queue*. A fluid queue is a two dimensional *Markov process* $\{(X_t, \phi_t) : t \in \mathbb{R}^+\}$, where

- $X_t \in \mathbb{R}^+$ is called "level" and represents the content of an infinite capacity fluid buffer at time t;

- $\{\phi_t : t \in \mathbb{R}^+\}$ is an *irreducible Markov process*, which represents "state" at time t and which takes values in some finite state space; the process regulates the evolution of the buffer content. We denote by T the infinitesimal transition generator of ϕ_t.

During time intervals when ϕ_t is constant and equals i, the level X_t changes its value with rate r_i, which can take any real value. If $X_t = 0$ and the rate at time t is negative, then X_t remains at zero:

$$\frac{dX_t}{dt} = \begin{cases} r_{\phi_t}, & \text{if} \quad X_t > 0, \\ \max(0, r_{\phi_t}), & \text{if} \quad X_t = 0. \end{cases} \tag{1.31}$$

Denote by $\underline{\pi}$ the stationary probability vector for the states of the fluid. It is defined by

$$\pi_i = \lim_{t\to\infty} \mathrm{P}[\,\phi(t) = i \,|\, \phi(0) = j\,], \quad \forall\, i, j. \tag{1.32}$$

$\underline{\pi}$ is the unique solution of

$$\begin{cases} \underline{\pi}T &= 0, \\ \underline{\pi}\mathbf{1} &= 1. \end{cases} \tag{1.33}$$

A fluid queue, where for state i the level evolves like a Brownian motion with drift μ_i and diffusion coefficient σ_i, is called a *Markov modulated Brownian motion* (Breuer [11]).

1.3 Mortality models

For hundreds of years now, actuaries have been using mortality tables as the basic tool to describe the mortality. However, a table is only an alternative when there is no suitable mathematical law of mortality available. According to Bowers et al. [10], the first mortality table was published by Edmund Halley in 1693 and the first mortality model by Abraham de Moivre in 1729. The majority of early models provide analytic expressions for the force of mortality termed a "law". For instance, the force of mortality of de Moivre's model is

$$\mu_x = (\omega - x)^{-1}, \qquad 0 \leq x < \omega.$$

The law of Gompertz (1825) is given by

$$\mu_x = Bc^x, \qquad B > 0,\, c > 1,\, x \geq 0.$$

Weibull's (1939) mortality force is

$$\mu_x = kx^n, \quad k > 0,\, n > 0,\, x \geq 0.$$

Both mortality tables and the approaches based on the force of mortality are "deterministic" and they tightly depend on data, a chosen pattern and chosen parameters. Seeking for a more flexible mortality structure, actuaries have developed many parametric and statistical models. A detailed description of these models is provided in Liu [42] and Pitacco [50]. One of the best known parametric mortality models was introduced in Lee and Carter [37]. The main difference of the Lee-Carter model from other parametric models is that it allows for uncertainty in forecasts of mortality. The model assumes that the central death rates introduced in Eq. (1.3) satisfy the equation

$$\ln m_x(t) = a_x + b_x k(t) + \epsilon_t, \tag{1.34}$$

where a_x, b_x are age-dependent parameters of the model, $k(t)$ describes the variation of the mortality level in time t; ϵ_t are independent Gaussian random variables for fixed t, with mean 0 and variance σ_ϵ^2. The age-dependence of the parameters a_x and b_x leads to different patterns of mortality for every age x and therefore makes it difficult to choose a proper compromise for $k(t)$. One way is to consider $k(t)$ as a time series and to apply an autoregressive model to describe its behavior. As a result, it becomes possible to estimate a confidence interval for the central rates $m_x(t)$. However, according to [42], many people believe that these confidence intervals are too narrow, which implies an underestimation of the risk of more extreme events. Therefore, a chosen pattern and data play a significant role in the mortality forecasting. All this implies that if the pattern changes in the future the model will not work. The cases when the pattern changes are numerous.

It has been shown by many studies that the life expectancy has steadily increased over time and human mortality rates keep decreasing (Oeppen and Vaupel [47]). In Liu [42] the Swedish male population is taken as an example in order to demonstrate the change of mortality rates. The author observes an overall increase in the most probable age of death, an overall decrease in mortality rates at all ages and an overall increase in the life expectancy. Another interesting observation is that the mortality does not change evenly for all the ages. Specifically, the mortality decreases more for young adult ages than for older ages. We verify this on the Belgium population of mixed genders (the data source is [62]) from 1902 to 2007. In Fig. 1.2 and Fig. 1.3 we show the evolution of Belgian mortality rates and the average life expectancy taken from the available mortality tables. The figures show that the conclusions obtained in Liu [42] for Swedish male population also hold for the Belgian population.

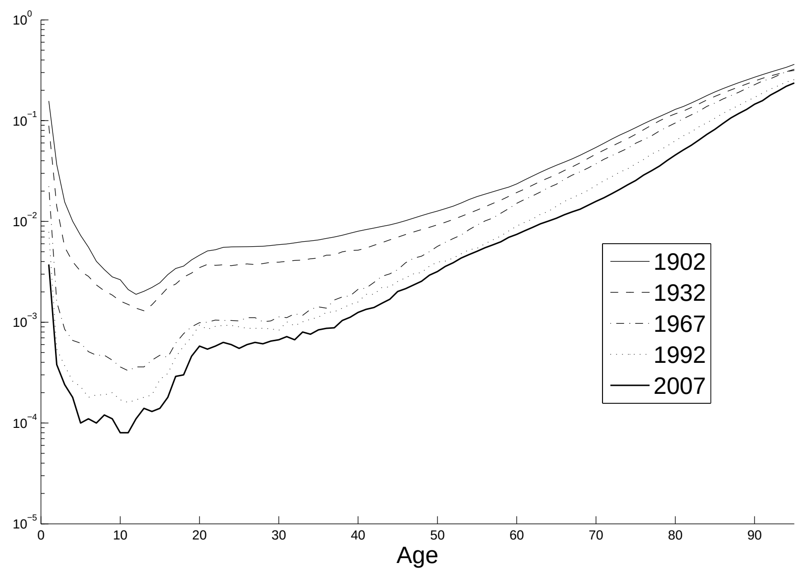

Figure 1.2: Log mortality rates, $\log q_x$, Belgium, 1902 – 2007

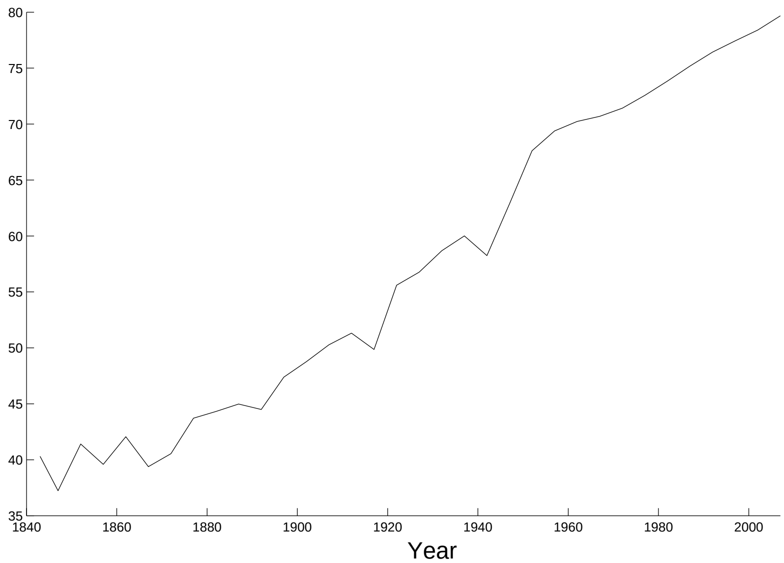

Figure 1.3: Average life expectancy, Belgium, 1840 – 2007

1.4 Phase-type methods in insurance

As we have already mentioned, phase-type methods are convenient to describe objects that can be characterized by different states and are widely used in numerous research areas. In insurance, the methods were used already in 1969 (Hoem [31]). The phase-type approach arises when the benefits are dependent on the status of an insured individual, which is clearly the case for pension, life and disability insurance. The approach allows for a natural interpretation, and often provides convenient matrix expressions for basic actuarial quantities. A good survey of actuarial calculations using Markovian models is presented in Jones [35]; in this section we present only essential aspects.

In individual life and disability insurance often a two- or three-state model is used to model an insurance policy. These models are schematically represented in Fig. 1.4. For a life annuity described by the two-state model, it is assumed that benefits are paid while the individual is in state "Alive" and cease upon transition to state "Dead". The three-state model can be used to describe a disability income policy. Here, premiums are payable while the individual is in state "Alive", and benefits are payable while the individual is in state "Disabled". The calculation of actuarial values for simple state Markov models is given in Bowers et al. [10].

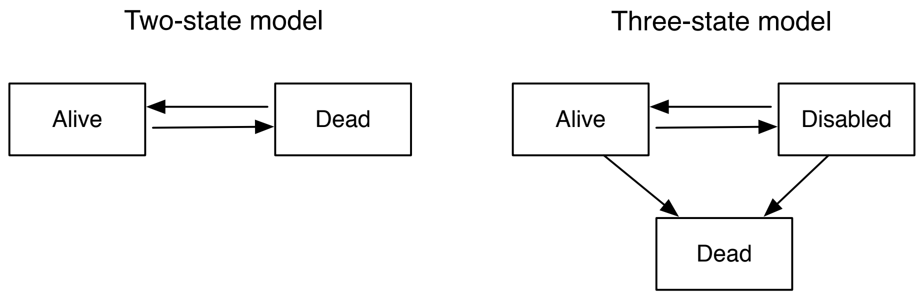

Figure 1.4: Simple state space Markov models in insurance

Markov chains are widely applied to describe the dynamic of a population, for example, of a pension fund. In these applications an insured individual is assumed to be in one of the predefined sets of states with given transition probabilities. According to Mettler [44], pension plan participants can be divided into the following subcategories within a pension fund : active employees, disabled benefit recipients, retired benefit recipients, deceased beneficiaries and resigned beneficiaries. In this case, Eq. (1.20) and Eq. (1.24) represent the dynamics of the population of pension plan participants. The works of Bertschi et al. [9], Janssen and Manca [34] and Mettler [44] provide an extended application of this method for profit tests in pension insurance.

In health care, there have been many studies on the estimation of medical costs over a fixed period of time, where the health of an individual is modeled by a Markov chain with pre-defined and/or observable states. Most of these studies use the Markovian approach to estimate the mean of the total cost based on available medical records. In Gardiner et al.[25] the authors obtain expected net present values of total costs, in Castelli et al.[17] a parametric assumption is introduced to estimate the mean total cost for a patient with a cancer disease and to perform a cost-efficiency analysis. A recent work by Zhao and Zhou [67] provides an accurate estimation of medical costs data by using copula methods.

In economics, the force of interest can be modeled as function of an underlying Markov chain, the states of which represent different states of the economical environment, see Norberg [46].

In insurance ruin theory, see Asmussen and Albrecher [4], *fluid queues* appear to be one of the most useful tools. In Stanford et al. [57], the authors use the connection between risk processes and fluid queues to obtain efficient computational algorithms to determine the probability of ruin prior to an Erlangian horizon, assuming that the sizes of the claims are phase-type distributed. In Badescu et al. [6] the fluid queue approach is implemented to determine the Laplace transform of the time until ruin for a general risk model. Badescu and Landriault [7] develop a recursive algorithm for the calculation of the moments of the discounted dividend payments before ruin, assuming a multi-threshold dividend structure. Here, the authors establish a connection between an insurer's surplus process and a corresponding fluid queue.

Markov modulated Brownian motion is widely applied in both ruin theory (for the Laplace transform of the time to ruin see Breuer [11]) and mathematical finance (for option pricing see Elliot and Swishchuk [23]).

Chapter 2

Phase-type aging model

We begin this chapter by Section 2.1, where we introduce the phase-type aging model and its properties. One of the first logical questions that arises when a new tool appears in an environment is "how applicable is it and what type of problems does it help to solve?". In the current framework it means that before applying the PH-aging model to complex actuarial problems, we find it necessary to verify how suitable it is with respect to the basic ones. Specifically, in Section 2.2 we apply the phase-type aging model to compute various insurance premiums and to model the number of survivors in a population.

In Section 2.3 our intention is to show how one may establish a relation between the states of the PH-aging model and the age of an individual in different contexts. In Section 2.4 we investigate how the sudden change of health can be taken into account by the PH-aging model. Clearly, this is important for longevity risk estimations.

In the end of the chapter, in Section 2.5 we introduce the stochastic analogue of the PH-aging model, which was developed by Lin and Liu in [41].

2.1 Deterministic aging model

In [40], Lin and Liu use a finite-state continuous-time Markov process to model the hypothetical aging process, which is called *phase-type aging model* (below, "PH-aging model"). Each state represents a physiological age or a health state, and aging is described as a process of consecutive transitions from one health state to another. There is one absorbing state, and the transition from any other state to the absorbing state is interpreted as the death of the individual. The diagrammatic representation of the proposed aging process is shown in Fig. 2.1. Here, the system has n phases with the transition rates λ_i, $i = 1, \ldots, n-1$ and the q_is are the transition rates to the absorbing state.

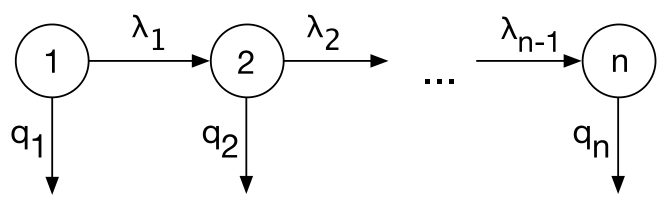

Figure 2.1: Phase-type aging process

All newborns start from the phase 1, therefore the initial probability vector $(\alpha_0, \underline{\alpha})$ is such that $\alpha_0 = 0$ and $\underline{\alpha} = [\,1\,0\,\ldots\,0\,]$.

For this model, the time to reach the absorbing state has the meaning of the age at death. The generator of this Markov process is given by

$$Q = \begin{bmatrix} \Lambda & \underline{q}^{\mathsf{T}} \\ \underline{0} & 0 \end{bmatrix} \tag{2.1}$$

where $\underline{q} = [\,q_1 \,\ldots\, q_n\,]$ and

$$\Lambda = \begin{pmatrix} -(\lambda_1 + q_1) & \lambda_1 & 0 & \ldots & 0 \\ 0 & -(\lambda_2 + q_2) & \lambda_2 & \ldots & 0 \\ 0 & 0 & -(\lambda_3 + q_3) & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & -q_n \end{pmatrix}, \tag{2.2}$$

and the time of death follows the phase-type distribution with parameters $(\underline{\alpha}, \Lambda)$. The expression $(\underline{\alpha} e^{\Lambda t})_i$ is the probability to survive for t years and to be in the phase i at time t. We denote by $\underline{\tau}_x$ the health state distribution of an individual who reaches age x. According to the model, it is given by

$$\underline{\tau}_x = \frac{\underline{\alpha} e^{\Lambda x}}{\underline{\alpha} e^{\Lambda x} \mathbf{1}}. \tag{2.3}$$

Obviously, the time until death of an individual of age x follows the phase-type distribution with parameters $(\underline{\tau}_x, \Lambda)$ (Latouche and Ramaswami [36]). The survival probability that such an individual survives for t units of time is given by

$$S_x(t) = \underline{\tau}_x e^{\Lambda t}\mathbf{1}. \tag{2.4}$$

The survival function $S(t)$ introduced in Eq. (1.26) has the same meaning as the survival probability ${}_tp_0$, and, in general, $S_x(t)$ is identical to ${}_tp_x$ from Eq. (1.1).

The authors define a structure of the parameters q_i, $i = 1, \ldots, n$ and λ_i, $i = 1, \ldots, n-1$, based on the physiological properties of population they observed. A detailed explanation is given in Lin and Liu [40], here we only give a brief outline.

$$\lambda_i = \begin{cases} \lambda_i, & i \leq k \\ \lambda, & \text{otherwise} \end{cases} \quad \text{and} \quad q_i = \begin{cases} b + a + i^p q, & i_1 \leq i \leq i_2 \\ b + i^p q, & \text{otherwise,} \end{cases}$$

where k, i_1, i_2, a , b, p, q are parameters of the model. The meaning is that the authors mark out a developmental period of the phases for very young ages, characterized by a different value of λ_i for every phase, and also a period of increased accident probability, for juveniles, corresponding to the range of phases i_1 to i_2, with higher death rates.

The model potentially consists of a large number of parameters, however, the authors show that 9 to 13 parameters are enough to give a good approximation of mortality rates. By comparison, the usual number of parameters of the Lee-Carter model is around 300. The parameters in Λ are estimated by minimizing the functional

$$\phi = \sum_{x=0}^{w} \left(q_x - \hat{q}_x\right)^2 S(x), \tag{2.5}$$

where q_x and $S(x)$ are the observed death rate and survival probability at age x, $\hat{q}_x$ is the corresponding model value for q_x and w is the maximal age available in the observations. Here, $\hat{q}_x$ can be expressed in terms of survival functions for ages x and $x+1$

$$\hat{q}_x = \frac{\underline{\alpha} e^{\Lambda x}\mathbf{1} - \underline{\alpha} e^{\Lambda(x+1)}\mathbf{1}}{\underline{\alpha} e^{\Lambda x}\mathbf{1}} = 1 - \underline{\tau}_x e^{\Lambda}\mathbf{1}.$$

We illustrate in Fig. 2.2 the result of the fitting procedure obtained in Lin and Liu [40]. The solid line represents $\log(10^3 \hat{q}_x)$ obtained with the phase-type model, the crosses are the mortality rates taken from SW1911M. The parameters of Λ are

$$n = 200, \quad \lambda = 2.3707, \quad b = 9.0987 \cdot 10^{-4}, \quad a = 2.8939 \cdot 10^{-3}, \tag{2.6}$$

$$q = 1.8872 \cdot 10^{-15}, \quad p = 6, \quad i_1 = 33, \quad i_2 = 70, \tag{2.7}$$

the length of the developmental period $k = 4$ and

$$q_1 = 0.1671, \quad q_2 = 0.0097, \quad q_3 = 0.0003, \quad q_4 = 0.0149, \tag{2.8}$$

$$\lambda_1 = 1.7958, \quad \lambda_2 = 0.5543, \quad \lambda_3 = 3.5061, \quad \lambda_4 = 0.6535. \tag{2.9}$$

In this thesis, in all figures with "SW1911M" in the caption the parameters of the underlying PH-aging model are given by Eq. (2.6)-(2.9).

For the questions of this thesis we are interested only in the survival probability, not in the mortality rates. By performing the minimization procedure (2.5) on the same data we have found a good approximation of the survival function, which we demonstrate in Fig. 2.3. The crosses represent the survival function computed from the mortality table and the solid line stands for the survival function of the PH-aging model.

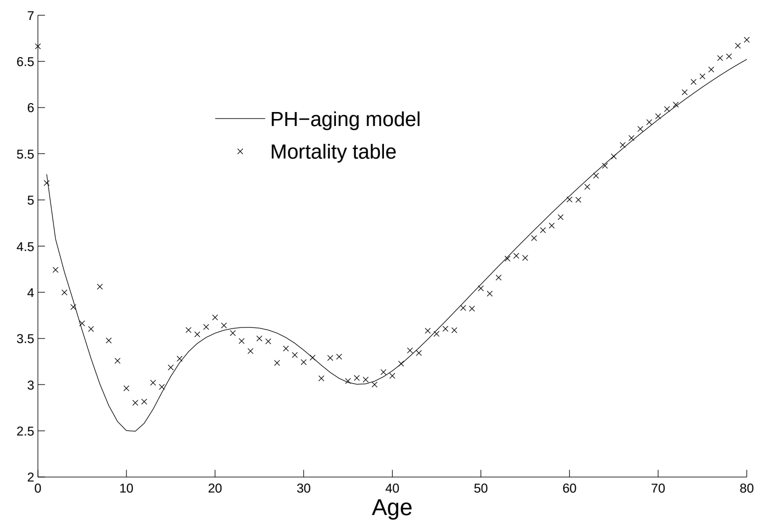

Figure 2.2: Mortality rates, $\log(10^3 q_x)$, SW1911M

Another way to find the parameters of Λ is to use the fitting algorithm suggested in Asmussen [5] for phase-type distributions. The method uses the expectation-maximization algorithm to approximate a sample density with a phase-type distribution. The implementation of this algorithm is available as a software tool with a full userguide paper written by Olsson [48]. As described in the userguide, the program looks for estimates of the initial vector $\underline{\alpha}$ and the generator Λ starting from

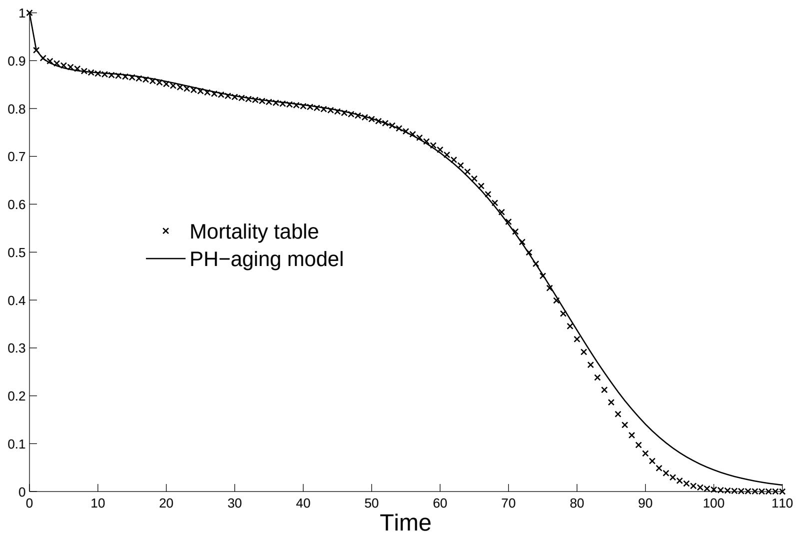

Figure 2.3: Survival function, $S(t)$, SW1911M

some initial values $(\underline{\alpha}^{(0)}, \Lambda^{(0)})$ which can be either given by the user or can be randomly generated by the program. At every iteration the program produces a new estimation of these parameters such that its likelihood function increases. Namely, the estimation $(\underline{\alpha}^{(k+1)}, \Lambda^{(k+1)})$ is such that

$$L(\underline{\alpha}^{(k)}, \Lambda^{(k)}; \underline{y}) \leq L(\underline{\alpha}^{(k+1)}, \Lambda^{(k+1)}; \underline{y})$$

where $\underline{y}$ is the sample to be fitted and $L(\underline{\alpha}, \Lambda; \underline{y}) = -\prod_{i=1}^{n} \underline{\alpha} e^{\Lambda y_i} \Lambda \mathbf{1}$. We have experimented with this method and we have not obtained a satisfactory fit of the data. The reason is simple: the algorithm of Asmussen works best with a small number of states; in order to obtain a good fitting of the mortality rates we need about 200 phases as explained in Lin and Liu [40]. The best approximation that we obtained with about 30 states is illustrated in Fig. 2.4, where we give $1 - {}_tp_0$, the probability to die at age t or before for a newborn individual.

There are also other methods of fitting phase-type distributions, see, for example, Horvath and Telek [32].

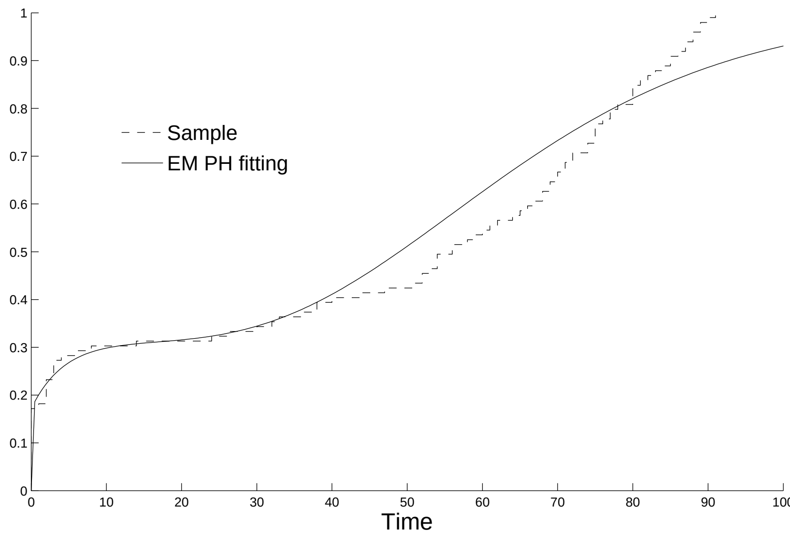

Figure 2.4: Fitting of $1 - S(t)$, SW1911M

2.2 Towards actuarial applications

In order to illustrate the PH-aging model we consider two basic actuarial problems: pricing and a population modeling.

Pricing. *Lifetime annuities.* We begin by applying the traditional balance approach to calculate the values of lifetime annuities for the health states of the PH-aging model.

Let us introduce the probability ${}_k p^{(j)}$ to survive k years for an individual in the health state j at time 0, it is given by

$$ {}_k p^{(j)} = \underline{\alpha}^{(j)} e^{\Lambda k} \mathbf{1}, \qquad \text{for } k \geq 0, \tag{2.10}$$

where, Λ is defined by Eq. (2.2) and $\underline{\alpha}^{(j)}$ is a row-vector of size n with

$$ \alpha_j^{(j)} = 1, \qquad \alpha_i^{(j)} = 0 \quad \text{for } i \neq j. \tag{2.11}$$

The value of the lifetime annuity for an individual in the health state j is then defined by

$$\ddot{a}^{(j)} = \sum_{k=0}^{\infty} v^k {}_k p^{(j)} \tag{2.12}$$

$$= \underline{\alpha}^{(j)} \sum_{k=0}^{\infty} v^k e^{\Lambda k} \mathbf{1} \ = \underline{\alpha}^{(j)} (I - v e^{\Lambda})^{-1} \mathbf{1}. \tag{2.13}$$

We consider a lifetime annuity with one unit payment per year at the beginning of each year. By definition, for an individual at calendar age x, this annuity is defined in Eq. (1.12) and can be rewritten as

$$\ddot{a}_x = \sum_{k=0}^{\infty} v^k S_x(k), \tag{2.14}$$

where v is the discount factor, $S_x(k)$ is the k-years survival probability for an individual aged x, defined in Eq. (2.4). With $\underline{\tau}_x$ defined by Eq. (2.3) we easily verify that $\ddot{a}_x$ is expressed as

$$\ddot{a}_x = \sum_{j=1}^{n} \left(\underline{\tau}_x\right)_j \ddot{a}^{(j)} \ = \underline{\tau}_x (I - v e^{\Lambda})^{-1} \mathbf{1}. \tag{2.15}$$

The equations (2.4, 2.12, 2.15) allow us to use conditioning arguments based on the calendar age or on the physiological age with equal ease.

We use the fact that $\sum_{k=0}^{t-1} v^k e^{\Lambda k} = (I - v^t e^{\Lambda t})(I - v e^{\Lambda})^{-1}$ to determine the term annuity for t years defined in Eq. (1.13)

$$\ddot{a}_{x:\overline{t}|} = \sum_{k=0}^{t-1} v^k S_x(k) = \underline{\tau}_x (I - v^t e^{\Lambda t})(I - v e^{\Lambda})^{-1} \mathbf{1}. \tag{2.16}$$

For comparison purpose, we represent in Fig. 2.5 the lifetime annuities $\ddot{a}_x$ for ages 1–65 calculated from (2.15) as a solid line and the values directly calculated from the mortality table as crosses. We observe that the fit is quite good, although there is a slight difference for ages 1–10 and for ages over 50. To explain the difference, we refer to Fig. 2.3 for the corresponding survival functions, $S(t)$, for a newborn individual. By looking at the graph we observe that the fit is very good for t in the range 20 to 80 years, however for earlier and later years one can notice slight discrepancies. The annuities $\ddot{a}_x$ are computed with the discount factor $v = 0.95$. Clearly, high powers of v converge to zero, which implies that for young individuals the fitting error after age 80 does not really matter. For individuals aged 20 to 50, the imperfect fit of the survival function for earlier ages does not matter and age 80 is still far, so that discounting will eliminate the fitting error; that is why we observe a good fit of $\ddot{a}_x$ for x in that range. For $x \leq 10$, however, the fitting imperfections of $S(t)$ in the first 20 years play a noticeable role. On the other hand, for individuals older than 50, age 80 is not that far, so the discounting does not eliminate completely the fitting error.

For practical purposes, the fit shown on Fig. 2.5 might be sufficiently good. If not, one would need to make a better approximation either by increasing the number of states in the model, or by minimizing a functional different from (2.5). For example, if younger/older ages are more important, in (2.5) one may increase the corresponding weights.

Another interesting observation that one can draw from Fig. 2.5 is that the annuities computed with the PH-aging model are always higher than the annuities directly calculated from the mortality table. Clearly, this is good for an insurer who wants to employ the PH-type approach, as it automatically allows one to incorporate some margin in the annuity values and, therefore, to be more cautious.

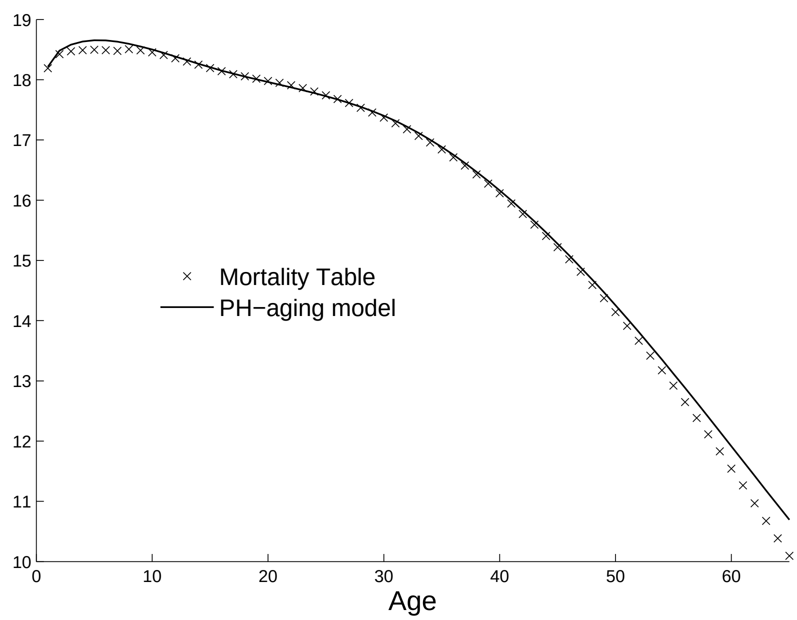

Figure 2.5: Lifetime annuities, $v = 0.95$, SW1911M

Term insurance. We consider a *term insurance* contract for n years and compute its price in discrete time. We notice that the single benefit premium for term insurance, given by Eq. (1.9) can be rewritten in terms of survival probabilities as

$$A^{1}_{x:\overline{n}|} = \sum_{i=0}^{n-1} v^{i+1}{}_{i}p_x - \sum_{i=1}^{n} v^{i}{}_{i}p_x = \sum_{i=0}^{n-1} v^{i+1} S_x(i) - \sum_{i=1}^{n} v^{i} S_x(i). \tag{2.17}$$

The net premium value is then given by

$$c_{x,n} = \frac{A^{1}_{x:\overline{n}|}}{\ddot{a}_{x:\overline{n}|}} = \frac{\sum_{i=0}^{n-1} v^{i+1} S_x(i) - \sum_{i=1}^{n} v^{i} S_x(i)}{\sum_{i=0}^{n-1} v^{i} S_x(i)}. \tag{2.18}$$

We use Eq. (2.16) to obtain

$$c_{x,n} = v - \frac{\underline{\tau}_x (I - v^{n+1} e^{\Lambda(n+1)})(I - v e^{\Lambda})^{-1}\mathbf{1} - 1}{\underline{\tau}_x (I - v^{n} e^{\Lambda n})(I - v e^{\Lambda})^{-1}\mathbf{1}},$$

which can be also written as

$$c_{x,n} = v - v\frac{\underline{\tau}_x (I - v^{n} e^{\Lambda n})(I - v e^{\Lambda})^{-1} e^{\Lambda}\mathbf{1}}{\underline{\tau}_x (I - v^{n} e^{\Lambda n})(I - v e^{\Lambda})^{-1}\mathbf{1}}. \tag{2.19}$$

In continuous time, we compute $c_{x,n}$ from the balance equation (Bowers [10])

$$c_{x,n} \int_0^n v^t S_x(t) dt = \int_0^n v^t f_x(t) dt, \tag{2.20}$$

where $f_x(t) = \underline{\tau}_x e^{\Lambda t} \underline{t}^{\top}$ is the density of the phase-type random variable with representation $(\underline{\tau}_x, \Lambda)$ (see Eq. (1.27) and Eq. (1.26)). By taking the integrals in Eq. (2.20) and $v = e^{-\delta}$, we obtain

$$c_{x,n} = \frac{\underline{\tau}_x \left(\Lambda - \delta I\right)^{-1} \left[e^{(\Lambda - \delta I -)n} - I\right] (-\Lambda)\mathbf{1}}{\underline{\tau}_x \left(\Lambda - \delta I\right)^{-1} \left[e^{(\Lambda - \delta I)n} - I\right] \mathbf{1}}. \tag{2.21}$$

We also compute the net premiums from the classical approach, with mortality rates from the table SW1911M. In Fig. 2.6, along with its classical values, we present the net premium values, computed with the continuous and discrete phase-type approach, for the ages from 1 to 60. One can see from the figure that all three methods give almost identical results.

Evolution of a population. Obviously, the mortality risks of a life insurance company strongly depend on the evolution of the population of its clients. In this

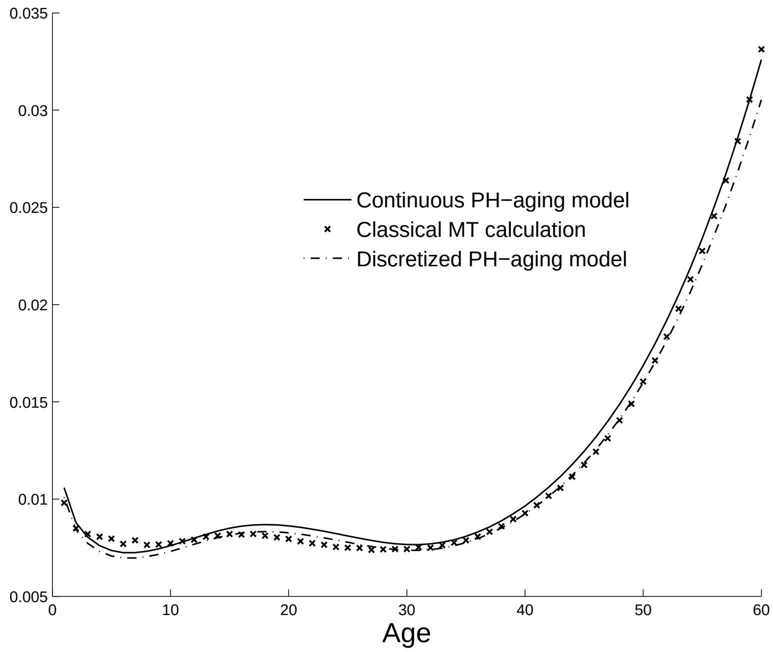

Figure 2.6: Premiums for term insurance, $c_{x,n}$
Parameters: ages 1-60, $n = 15$, $v = 0.943$, SW1911M

section we compare the phase-type aging model and existing classical models. To do this, we fix the age structure of a population, specify the total number of individuals in each age group and estimate the number of survivors in every year until a chosen time horizon.

Let $N_t(x)$ be the number of survivors of the age x at time t, $D_t(x)$ be the number of deaths of age x at time t. Thus, $N_{t_0}(x)$ represents a given population in the initial year of observation, t_0. The evolution of an age group can be represented by

$$N_{t+1}(x+1) = N_t(x) - D_t(x), \quad t = 1, ..., T.$$

We consider three methods to determine $N_{t_0+h}(x+h)$: the phase-type approach and two classical models that are based on the assumption of the binomial distribution of the number of deaths (see, for example, Pollard [51]).

Binomial simulation. Here, the number of deaths in year t has the binomial distribution, $D_t(x) \sim Bin(N_t(x), q_x)$, where $N_t(x)$ is the number of survivors at the beginning of the year and q_x is the mortality rate taken from the chosen mortality table. In this approach $N_{t_0+h}(x+h)$ is estimated through the simulation of numbers of deaths in every year. Thus, we perform a simulation of the random variable $\{D_t(x) : t = 1 \dots T\}$ many times and look at the average values of $N_{t_0+h}(x+h)$.

Theoretical binomial approach. The main assumption remains the same as in the previous model, $D_t(x) \sim Bin(N_t(x), q_x)$; and, symmetrically, the distribution of $N_{t+1}(x+1)$ is $Bin(N_t(x), 1-q_x)$, given $N_t(x)$. We define the generating function $F_t(\xi; x)$ such that

$$F_t(\xi; x) = \sum_{k=0}^{\infty} \xi^k \mathrm{P}\left[N_t(x) = k \right].$$

We use the definition of the binomial distribution to find $F_{t+1}(\xi, x+1)$

$$\begin{aligned} F_{t+1}(\xi; x+1) &= \sum_{k=0}^{\infty} \xi^k \mathrm{P}\left[\ N_{t+1}(x+1) = k\ \right] \\ &= \sum_{k=0}^{\infty} \sum_{i} \mathrm{P}\left[\ N_t(x) = i\ \right] \mathrm{P}\left[\ N_{t+1}(x+1) = k \mid N_t(x) = i\ \right] \xi^k \\ &= \sum_{i} \mathrm{P}\left[\ N_t(x) = i\ \right] \left[q_x + (1-q_x)\xi\right]^i \\ &= F_t(q_x + \xi p_x; x), \end{aligned} \tag{2.22}$$

where q_x and p_x are taken from the chosen mortality table. Therefore, in the first year after the year of initial observation, t_0, the generating function, is given by

$$F_{t_0+1}(\xi; x+1) = \left[q_x + (1-q_x)\xi\right]^{N_{t_0}(x)}. \tag{2.23}$$

Eq. (2.23) can be generalized as

$$F_{t_0+h}(\xi; x+h) = \left(1 - \prod_{i=0}^{h-1} p_{x+i} + \xi \prod_{i=0}^{h-1} p_{x+i}\right)^{N_{t_0}(x)}, \tag{2.24}$$

which we prove by induction. We assume that Eq. (2.24) holds for the year $t_0 + h$ and the age $x + h$. By Eq. (2.22),

$$\begin{aligned} F_{t_0+h+1}(\xi; x+h+1) &= F_{t_0+h}(q_{x+h} + \xi p_{x+h}; x+h) \\ &= \left(\left(1 - \prod_{i=0}^{h-1} p_{x+i}\right) + (1-p_{x+h}) \prod_{i=0}^{h-1} p_{x+i} + \xi \prod_{i=0}^{h} p_{x+i}\right)^{N_{t_0}(x)} \\ &= \left(1 - \prod_{i=0}^{h} p_{x+i} + \xi \prod_{i=0}^{h} p_{x+i}\right)^{N_{t_0}(x)}. \end{aligned}$$

This allows us to conclude that the distribution of survivors after h years is

$$N_{t_0+h}(x+h) \sim Bin(N_{t_0}(x), p_x \dots p_{x+h-1}),$$

or, equivalently, $N_{t_0+h}(x+h) \sim Bin(N_{t_0}(x), {}_h p_x)$.

Phase-type aging model. We assume that the individuals are independent and all follow the same PH-aging model. This implies that

$$D_{t_0+h}(x+h) \sim Bin(N_{t_0}(x), \underline{\tau_x} e^{\Lambda(h-1)}\mathbf{1} - \underline{\tau_x} e^{\Lambda h}\mathbf{1}),$$

where the second parameter represents the probability of dying between years $h-1$ and h. Thus, the expected number of deaths is given by

$$E[D_{t_0+h}(x+h)] = N_{t_0}(x)(\underline{\tau}_x e^{\Lambda(h-1)}\mathbf{1} - \underline{\tau}_x e^{\Lambda h}\mathbf{1}).$$

The comparison of the three methods is shown in Fig. 2.7. As initial population, we considered 10000 individuals of male gender with uniformly distributed age-structure; the cohort mortality table is SW1911M. We performed the simulation of both binomial methods and estimated the expected number of survivors in future years. The same quantity was calculated using the phase-type approach. In the figure we plot the ratio of the numbers obtained from the binomial methods over those from the phase-type method. As it can be seen from the figure the maximum difference does not exceed 0.2%.

2.3 Characterization of health states

In this section our focus is to determine possible health states of an individual in the phase-type aging model using available data. Clearly, the available data about the individual contains the information about his/her age. In Sec. 2.1, we provided a formula to compute the distribution of health states for a given age x, which can be considered as a first approximation of a possible health state. It is given by

$$\underline{\tau}_x = \frac{\underline{\alpha} e^{\Lambda x}}{\underline{\alpha} e^{\Lambda x}\mathbf{1}},$$

where $\underline{\alpha}$ is the health distribution at birth.

Also, the available data may contain either direct information about the health of the individual, coming, for example, from expert opinions, or indirect objective information like, for example, health treatment costs consumed in the past years.

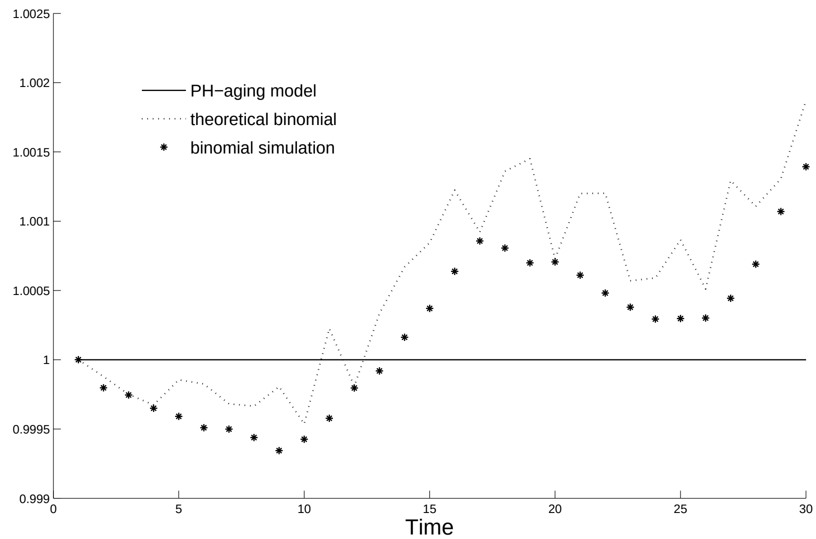

Figure 2.7: Binomial methods vs PH-aging model: ratio of survivors, SW1911M.

Direct information. As mentioned in Brockett [13], often expert opinions about the future life expectancy of an individual are available, and force the recalculation of the mortality table for this individual. Let us assume that the expected lifetime, $\bar{e}$, is determined. Denote by $e^{(i)}$ the expected number of the remaining years of life for each health state i. For an individual in health state i the probability to live exactly k years is $\underline{\alpha}^{(i)}e^{\Lambda k}(I-e^{\Lambda})\mathbf{1}$, where $\underline{\alpha}^{(i)}$ is a row-vector of size n defined by Eq. (2.11). Therefore, $e^{(i)}$ is given by

$$e^{(i)} = \underline{\alpha}^{(i)} \sum_{k=0}^{\infty} k e^{\Lambda k}(I-e^{\Lambda})\mathbf{1}.$$

If we estimate $\bar{e}$, one might determine the most likely health state i^*, because there is always i^* such that $e^{(i^*+1)} \leq \bar{e} \leq e^{(i^*)}$, which suggests that the health state of the individual is going to be somewhere around i^*.

Indirect information. Suppose that one associates to an individual some characterization, which is indirectly related to his/her age. For instance, this characterization can represent health care expenses at a given age. Specifically, assume that an individual aged x can be characterized by the realization of the discrete random variable A_x. We also assume that the average values of A_x are known for every age:

$E[A_x] = a_x,\ x = 1, ..., x_{max}$. This information allows us to determine the average values of this random variable for every health state, $\underline{b} = (b^{(1)}, ..., b^{(n)})$. We compose the linear system

$$\mathcal{T}\underline{b}^{\mathrm{T}} = \mathcal{J}, \tag{2.25}$$

where $\mathcal{T} = [\underline{\tau}_x]_{x=1,..,x_{max}}$, $\mathcal{J} = [a_x]_{x=1,..,x_{max}}$, and where $\underline{\tau}_x\underline{b}^{\mathrm{T}}$ represents the average value of A_x, computed from the PH-aging model. In order to determine $\underline{b}$ we minimize

$$\min_{\underline{b}} \|\mathcal{T}\underline{b}^{\mathrm{T}} - \mathcal{J}\|_2^2. \tag{2.26}$$

From Sec.2.1 we know that the number of health states in the PH-aging model can be quite large, specifically, $n = 200$. Therefore, linear system (2.25) becomes underdetermined as the number of ages x_{max} can not be much more than 100. Thus, in order to solve problem (2.26) we find it useful to have a constraint. Obviously, the choice of the constraint should depend on the nature of the random variable A_x. If A_x is related to the health care cost of the individual, one logical constraint is given by

$$0 \leq b^{(i)} \leq b^{(i+1)} \leq \max_x(a_x), \tag{2.27}$$

which implies that the health care costs do not decrease with the health state and may not exceed the maximal value given by the data. We observe that the optimization problem with the indicated constraint guarantees a piecewise constant step structure of the costs for the health states. This type of solution allows one to determine a range of possible health states for a given cost value. Let us consider an example. We take the data on average health care costs for ages provided in Table 2.1, taken from [61], to obtain the costs for the health states in the problem (2.26) with constraint (2.27). The results are shown in Fig. 2.8. We observe that the results provide a good fit: in Fig. 2.9 we present the average costs for different ages x computed from the phase-type model, $\underline{\tau}_x\underline{b}^{\mathrm{T}}$, next to the average costs given by the data, a_x. As one can see in Fig. 2.8, if an individual spends, for example, \$8000 per year for his/her health treatment, the possible health states are 163-200 (conventionally); if \$2000, then 51-103. In Fig. 2.10 we demonstrate the distribution of health states at ages 50, 60 and 70, given by Eq. (2.3) together with the possible states given the cost of \$2000 and of \$8000. We find from Eq. (2.3) that, if we take an individual aged 60, then the probability mass is concentrated for health states 90-182. If, in addition, we know that the treatment cost equals \$2000, then we find that the possible health state lays in the interval 90-103; if it equals \$8000 – in the interval 163-182.

Table 2.1: Average Annual Health Care Costs by Age Group

Age group	Average Cost
0	$8318
1-4	$1285
5-9	$1095
10-14	$1202
15-19	$1667
20-24	$1967
25-29	$2371
30-34	$2484
35-39	$2541
40-44	$2808
45-49	$3575
50-54	$4037
55-59	$4407
60-64	$4999
65-69	$6028
70-74	$7612
75-79	$11195
80-84	$15150
85-89	$21650
90+	$8631

2.4 Longevity risk estimation tools

As we have already mentioned, an unexpected change of health causes additional uncertainties about the remaining lifetime of individuals and is called a longevity risk. The question is about how such unexpected changes may be modeled. We make a distinction between two different natural reasons. The first one is related to internal factors, that is, properties of the human body which are defined by genetics and by personal habits. The second is related to external factors that externally affects the life of individual, like economics, medical service, scientific developments, etc. The incorporation of the health status through the PH-aging model allows us to separately consider the two indicated reasons.

Internal factors. This effect is obtained through a change of the initial health state distribution. For instance, we may assume that an individual of age x con-

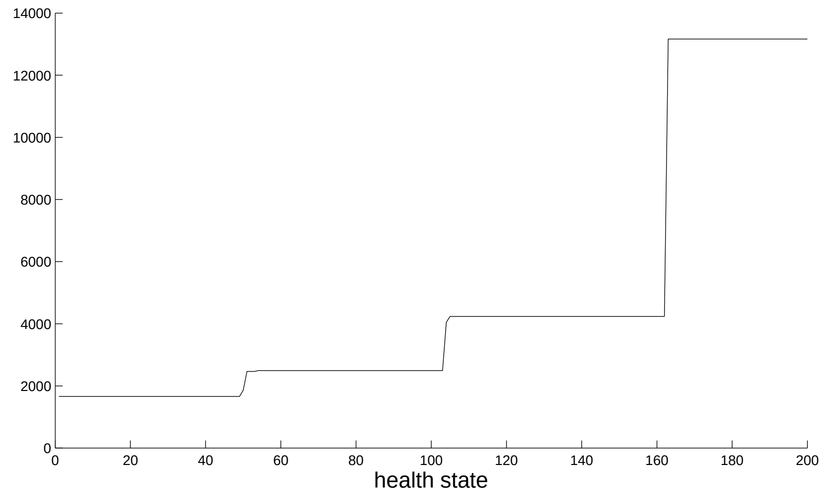

Figure 2.8: Health care costs for health states
SW1911M, Table 2.1.

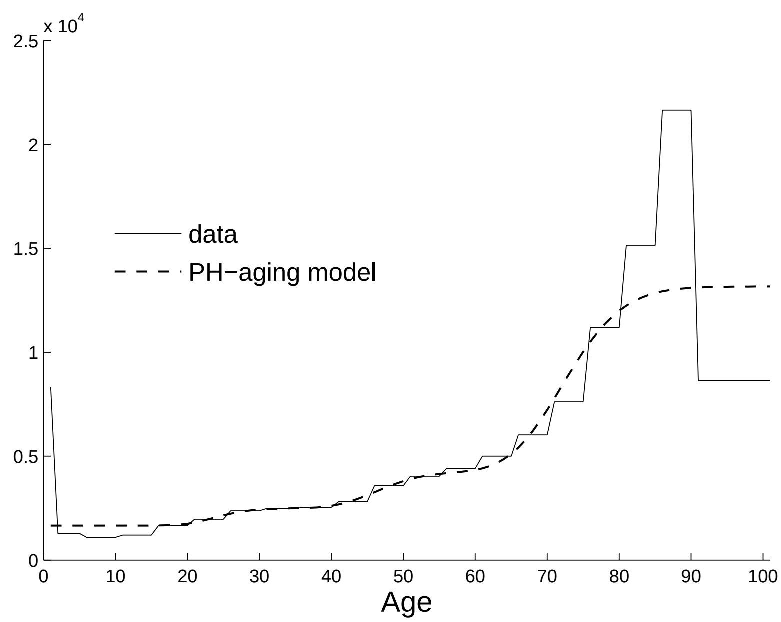

Figure 2.9: Health care costs for ages
SW1911M, Table 2.1.

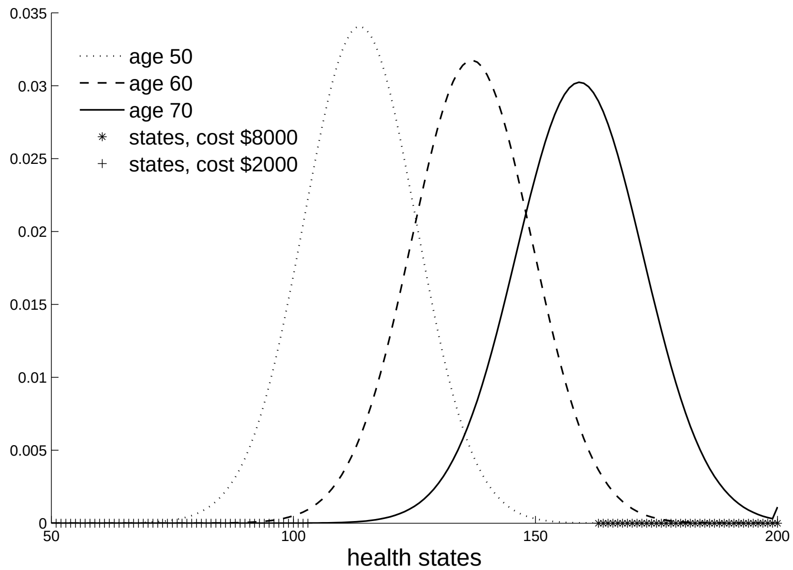

Figure 2.10: Characterization of health states
SW1911M, Table 2.1.

sidered at time 0 has the same health as if he was γ years younger. In this case, the initial health distribution would be given by $\underline{\tau}^*_x = \underline{\tau}_{x-\gamma}$. Alternatively, we might chose to shift the distribution over the phases, so that the density profile is the same but the health state is younger, and define $(\underline{\tau}^*_x)_i = (\underline{\tau}_x)_{i+\delta}$, for $1 \le i \le n - \delta$, where n is the number of health states; unless x is very small, the missing mass $(\underline{\tau}_x)_1 + \cdots + (\underline{\tau}_x)_{\delta-1}$ is likely to be small and may be assigned to $\underline{\tau}^*_x$ in any reasonable way. The new survival function is

$$S^*_x(t) = \underline{\tau}^*_x e^{\Lambda t}\mathbf{1}. \tag{2.28}$$

This idea will be applied in Section 6.3 to estimate the impact of *early retirement*.

Environmental changes. Here, we assume that at some time the mortality rates of the whole population become lower. Specifically, at time K in the future the phase-type aging matrix Λ and the vector of mortality rates are perturbed as follows

$$\tilde{\Lambda} = \Lambda + \varepsilon D(\underline{q}), \quad \tilde{q} = (1 - \varepsilon)\underline{q}, \tag{2.29}$$

where ε is a positive scalar assumed to be small and $D(\underline{q})$ is the diagonal matrix with vector $\underline{q}$ on the diagonal. The effect is to reduce mortality, but to keep physiological evolution unchanged. The new survival function is

$$S_x^{\varepsilon}(t) = S_x(t), \text{ for } t < K, \quad S_x^{\varepsilon}(t) = \underline{\tau}_x e^{\Lambda K} e^{\tilde{\Lambda}(t-K)} \mathbf{1}, \text{ for } t \geq K. \tag{2.30}$$

This method will be used Section 7.5 to study the effect of an increasing lifetime spent in bad health states. Also, in this case, instead of the single perturbation coefficient ε, one can apply different coefficients $\varepsilon_1, ..., \varepsilon_n$ for each of the health states.

2.5 Stochastic aging model

In Lin and Liu [41], the authors incorporate stochasticity into the survival function $S(t)$, defined in (1.26), by introducing a time-changed Markov process. Denote by J_t the state at time t for the phase process of Section 2.1. Let γ_t be a nondecreasing continuous-time stochastic process. Denote by Z_t the health state at time t of a given individual aged x at time 0. It is defined as the subordinated Markov process

$$Z_t = J_{\gamma_t}. \tag{2.31}$$

In other words, it is assumed that the phase-type aging process is influenced by the time-change process γ_t. Under this stochastic mortality modeling framework, the survival function for a newborn is now a stochastic process and it is given by

$$S^{\gamma}(t) = \underline{\alpha} e^{\Lambda \gamma_t} \mathbf{1}, \quad t > 0. \tag{2.32}$$

Similarly to Eq. (2.4), which corresponds to the deterministic PH-aging model, we define the survival function of an individual aged x as

$$S_x^{\gamma}(t) = \underline{\tau}_x e^{\Lambda \gamma_t} \mathbf{1}, \quad t > 0. \tag{2.33}$$

The authors in [41] provide analytic expressions for the mean and the variance of $S^{\gamma}(t)$, assuming that γ_t is a subordinating *gamma process* with parameter $\mu = \theta = 1/k$ (Appendix A.3). This results will be useful in Section 3.2, where we deal with correlated cohorts.

Assume that the intensity matrix Λ has distinct eigenvalues, $-\lambda_1, ..., -\lambda_n$, and let $\underline{h}_1, ..., \underline{h}_n$ and $\underline{v}_1, ..., \underline{v}_n$ be their corresponding right and left eigenvectors such that $\underline{v}_i \underline{h}_i^{\mathrm{T}} = 1$. Define

$$\tilde{\Lambda} = -\sum_{i=1}^{n} \tilde{\lambda}_i \underline{h}_i^{\mathrm{T}} \underline{v}_i, \tag{2.34}$$

where

$$\tilde{\lambda}_i = \frac{1}{\mu} \ln(1 + \mu\lambda_i). \tag{2.35}$$

The expectation of $S^\gamma(t)$ is given by

$$E[S^\gamma(t)] = \underline{\alpha} e^{\tilde{\Lambda} t} \mathbf{1}, \quad t > 0. \tag{2.36}$$

Denote

$$H = (\underline{h}_1, ..., \underline{h}_n), \quad D = diag(-\lambda_1, ..., -\lambda_n)$$

and define $\zeta_1, ..., \zeta_{n^2}$ such that

$$\zeta_{i+j} = \lambda_i + \lambda_j, \quad i, j = 1, ..., n.$$

Define

$$\widetilde{D \oplus D} = diag(-\tilde{\zeta}_1, ..., -\tilde{\zeta}_{n^2}), \qquad \tilde{\zeta}_i = \frac{1}{\mu} \ln(1 + \mu\zeta_i)$$

and

$$\widetilde{\Lambda \oplus \Lambda} = (H \otimes H)(\widetilde{D \oplus D})(H \otimes H)^{-1}, \tag{2.37}$$

where the symbols $\otimes$ and $\oplus$ are Kronecker product and sum, respectively. By using the Kronecker operations on matrices (Appendix A.1), the authors show that

$$(E[S^\gamma(t)])^2 = (\underline{\alpha} \otimes \underline{\alpha}) \left[e^{(\tilde{\Lambda} \oplus \tilde{\Lambda})t} \right] (\mathbf{1} \otimes \mathbf{1}) \tag{2.38}$$

and

$$E\left[S^\gamma(t)^2\right] = (\underline{\alpha} \otimes \underline{\alpha}) \left[e^{(\widetilde{\Lambda \oplus \Lambda})t} \right] (\mathbf{1} \otimes \mathbf{1}). \tag{2.39}$$

The variance of $S^\gamma(t)$ is given by

$$Var[S^\gamma(t)] = (\underline{\alpha} \otimes \underline{\alpha}) \left[e^{(\widetilde{\Lambda \oplus \Lambda})t} - e^{(\tilde{\Lambda} \oplus \tilde{\Lambda})t} \right] (\mathbf{1} \otimes \mathbf{1}). \tag{2.40}$$

Chapter 3

Correlated cohorts

In this chapter we deal with correlated cohorts in stochastic mortality environment. We begin with our motivation in Section 3.1. In Section 3.2 we develop four mathematical methods of modeling correlated cohorts and provide their comparison. In order to show how these models can be applied, in Section 3.3 we consider the problem of hedging a fixed cohort of annuitants of lifetime pensions with the help of longevity bonds. We also present our numerical experiments and conclusions.

3.1 Why should cohorts be correlated?

A simple example of why cohorts are correlated is presented in Fig. 3.1. In the figure we demonstrate the survival functions of two cohorts of different age: cohort A and cohort B. On one side, at any given point of time the survival probabilities naturally differ from each other; on the other side, both cohorts live at the same time and, therefore, are subject to common shocks. These shocks may represent the invention of a new treatment, or an increase of the quality of life due to other reasons. The impact of these shocks may be different for each of the cohorts, but, obviously, it will have the same direction. The two natural questions that appear are how to introduce a correlation between cohorts and how to measure the impact of such correlation on the future profits of a life insurance company.

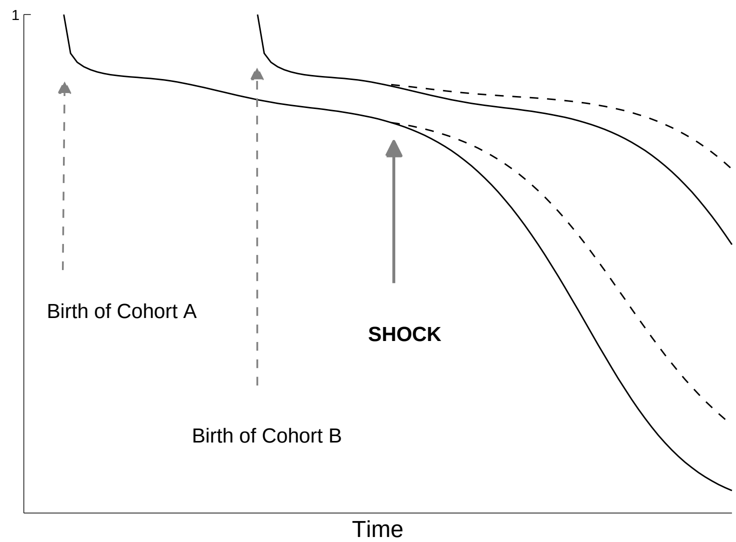

Figure 3.1: Survival functions of two correlated cohorts

3.2 Models of correlated cohorts

Consider two cohorts of different health: cohort A and cohort B, assumed to be the healthiest of the two. This may happen if they have different health state distribution. For instance, we may assume that the members of cohort A have age x and the members of cohort B are y years younger. Thus, their health state distribution is given by vectors $\underline{\tau}_x$ and $\underline{\tau}_{x-y}$, respectively (see Eq. (2.3)). In the deterministic PH-aging model, the corresponding survival functions are given by $S_x(t)$ and $S_{x-y}(t)$; in the stochastic aging model – by $S_x^\gamma(t)$ and $S_{x-y}^\gamma(t)$. According to Eq. (2.4), $S_x(t) = \underline{\tau}_x e^{\Lambda t}\mathbf{1}$, and, according to Eq. (2.33), $S_x^\gamma(t) = \underline{\tau}_x e^{\Lambda\gamma_t}\mathbf{1}$ for $t > 0$, where $\underline{\tau}_x$ is given by Eq. (2.3) and γ_t is a *gamma process* (Appendix A.3).

We develop below four methods of modeling correlated cohorts, that are based on the number of survivors in each of the cohorts. The first two deal with the stochastic survival function $S_x^\gamma(t)$, the last two deal with the deterministic survival function $S_x(t)$. Denote by $N^A(t)$ the number of survivors in the cohort A at time t, by $N^B(t)$ the number of survivors in the cohort B at time t. Assume also that $N^A(0) = N^B(0) = N_0$.

Method I. We adopt the classical Binomial assumption for the number of deaths in each year (Pollard [51]) in the stochastic mortality environment as follows

$$\begin{aligned} N^A(t) &\sim Bin(N_0, S_x^\gamma(t)), \\ N^B(t) &\sim Bin(N_0, S_{x-y}^\gamma(t)), \end{aligned} \tag{3.1}$$

where $S_z^\gamma(t)$ is a random variable that represents t year survival probability for an individual aged z and that depends on a gamma-process γ_t (see Eq. (2.33)). Here, $N^A(t)$ and $N^B(t)$ are correlated because they are controlled by the same stochastic process γ_t. In accordance with Eq. (2.36), the expectation of $N^A(t)$ is given by

$$\begin{aligned} E[N^A(t)] &= E_{\gamma_t} E[\, N^A(t) \,|\, \gamma_t \,] \\ &= E_{\gamma_t}[N_0 \underline{\mathcal{I}}_x e^{\Lambda\gamma_t}\mathbf{1}] = N_0 \underline{\mathcal{I}}_x e^{\tilde{\Lambda}t}\mathbf{1}, \end{aligned} \tag{3.2}$$

where $\tilde{\Lambda}$ is defined in Eq. (2.34). Due to the property given by Eq. (A.5)

$$\begin{aligned} (E[N^A(t)])^2 &= N_0^2 \underline{\mathcal{I}}_x e^{\tilde{\Lambda}t}\mathbf{1}\underline{\mathcal{I}}_x e^{\tilde{\Lambda}t}\mathbf{1} \\ &= N_0^2 (\underline{\mathcal{I}}_x \otimes \underline{\mathcal{I}}_x) e^{(\tilde{\Lambda}\oplus\tilde{\Lambda})t}(\mathbf{1}\otimes\mathbf{1}), \end{aligned} \tag{3.3}$$

where $\otimes$ and $\oplus$ are the Kronecker product and sum, respectively (Appendix A.1). We recall that $N^A(t)$ is Binomial to obtain

$$\begin{aligned} E\left[N^A(t)^2\right] &= E_{\gamma_t}\left[\, N_0 S_x^\gamma(t)(1 - S_x^\gamma(t)) + (N_0 S_x^\gamma(t))^2 \,|\, \gamma_t \,\right] \\ &= N_0 \underline{\mathcal{I}}_x e^{\tilde{\Lambda}t}\mathbf{1} + N_0(N_0 - 1)(\underline{\mathcal{I}}_x \otimes \underline{\mathcal{I}}_x) e^{(\widetilde{\Lambda\oplus\Lambda})t}(\mathbf{1}\otimes\mathbf{1}), \end{aligned} \tag{3.4}$$

where $\widetilde{\Lambda \oplus \Lambda}$ is defined in Eq. (2.37). To simplify the notations, we denote

$$B_t = \left[e^{(\widetilde{\Lambda\oplus\Lambda})t} - e^{(\tilde{\Lambda}\oplus\tilde{\Lambda})t}\right](\mathbf{1}\otimes\mathbf{1}). \tag{3.5}$$

Eq. (3.3) and Eq. (3.4) allow us to find

$$\begin{aligned} Var[N^A(t)] &= E\left[N^A(t)^2\right] - (E[N^A(t)])^2 \\ &= N_0 \underline{\mathcal{I}}_x e^{\tilde{\Lambda}t}\mathbf{1} - N_0(\underline{\mathcal{I}}_x \otimes \underline{\mathcal{I}}_x) e^{(\widetilde{\Lambda\oplus\Lambda})t}(\mathbf{1}\otimes\mathbf{1}) + N_0^2(\underline{\mathcal{I}}_x \otimes \underline{\mathcal{I}}_x) B_t \\ &= N_0(E[S_x^\gamma(t)] - E[S_x^\gamma(t)]^2) + N_0^2 Var[S_x^\gamma(t)]. \end{aligned} \tag{3.6}$$

If we replace x by $(x - y)$ in Eq. (3.2) and in Eq. (3.6), we obtain the expectation and the variance of $N^B(t)$.

In order to compute the covariance of $N^A(t)$ and $N^B(t)$ we find

$$\begin{aligned} E[N^A(t)N^B(t)] &= E_{\gamma_t}[\, E[\, N^A(t) \,|\, \gamma_t \,]\, E[\, N^B(t) \,|\, \gamma_t \,]\,] \\ &= E_{\gamma_t}[N_0^2 S_x^\gamma(t) S_{x-y}^\gamma(t) \,|\, \gamma_t \,] \\ &= N_0^2(\underline{\mathcal{I}}_x \otimes \underline{\mathcal{I}}_{x-y}) e^{(\widetilde{\Lambda\oplus\Lambda})t}(\mathbf{1}\otimes\mathbf{1}). \end{aligned} \tag{3.7}$$

We use Eq. (A.5) to note that

$$E[N^A(t)]E[N^B(t)] = N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})e^{(\tilde{\Lambda}\oplus\tilde{\Lambda})t}(\mathbf{1}\otimes\mathbf{1}),$$

and we obtain

$$\begin{aligned} Cov[N^A(t), N^B(t)] &= E[N^A(t)N^B(t)] - E[N^A(t)]E[N^B(t)] \\ &= N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})e^{(\widetilde{\Lambda\oplus\Lambda})t}(\mathbf{1}\otimes\mathbf{1}) - N_0^2\underline{\tau}_x e^{\tilde{\Lambda}t}\mathbf{1}\underline{\tau}_{x-y}e^{\tilde{\Lambda}t}\mathbf{1} \\ &= N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})e^{(\widetilde{\Lambda\oplus\Lambda})t}(\mathbf{1}\otimes\mathbf{1}) - N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})e^{(\tilde{\Lambda}\oplus\tilde{\Lambda})t}(\mathbf{1}\otimes\mathbf{1}) \\ &= N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})B_t = N_0^2 Cov[S_x^\gamma(t), S_{x-y}^\gamma(t)]. \end{aligned} \tag{3.8}$$

Method II. In method I we make a simplified assumption that the numbers of survivors behave as their expected values

$$\begin{aligned} N^A(t) &= N_0 S_x^\gamma(t) = N_0\underline{\tau}_x e^{\Lambda\gamma_t}\mathbf{1}, \\ N^B(t) &= N_0 S_{x-y}^\gamma(t) = N_0\underline{\tau}_{x-y}e^{\Lambda\gamma_t}\mathbf{1}. \end{aligned} \tag{3.9}$$

Due to the same argument as in method I, $N^A(t)$ and $N^B(t)$ are correlated. Moreover, they have the same expectations as in method I:

$$\begin{aligned} E[N^A(t)] &= N_0 E\underline{\tau}_x e^{\Lambda\gamma_t}\mathbf{1} = N_0\underline{\tau}_x e^{\tilde{\Lambda}t}\mathbf{1}, \\ E[N^B(t)] &= N_0 E\underline{\tau}_{x-y}e^{\Lambda\gamma_t}\mathbf{1} = N_0\underline{\tau}_{x-y}e^{\tilde{\Lambda}t}\mathbf{1}. \end{aligned} \tag{3.10}$$

In accordance with Eq. (2.40) and the introduced notation, the variances are

$$\begin{aligned} Var[N^A(t)] &= N_0^2(\underline{\tau}_x \otimes \underline{\tau}_x)B_t = N_0^2 Var[S_x^\gamma(t)], \\ Var[N^B(t)] &= N_0^2(\underline{\tau}_{x-y} \otimes \underline{\tau}_{x-y})B_t = N_0^2 Var[S_{x-y}^\gamma(t)]. \end{aligned} \tag{3.11}$$

If we compare this equation with Eq. (3.6), we conclude that in method I the impact of the Binomial assumption on the variance exactly equals $N_0(E[S_x^\gamma(t)] - E[S_x^\gamma(t)]^2)$ for $N^A(t)$ and $N_0(E[S_{x-y}^\gamma(t)] - E[S_{x-y}^\gamma(t)]^2)$ for $N^B(t)$.

By using Eq. (3.7), Eq. (3.8) and Eq. (3.10), we obtain

$$Cov[N^A(t), N^B(t)] = N_0^2(\underline{\tau}_x \otimes \underline{\tau}_{x-y})B_t = N_0^2 Cov[S_x^\gamma(t), S_{x-y}^\gamma(t)], \tag{3.12}$$

which is the same as in method I. Thus, from Eq. (3.10), (3.11) and (3.12) we observe that the Binomial assumption of method I increases only the variance of $N^A(t)$ and $N^B(t)$; the expectations and the covariance remain the same.

Method III. We keep the classical Binomial assumption for the number of deaths in each year, but we remove the stochasticity in the survival probability. Specifically, we suggest that

$$N^A(t) \sim Bin(N_0, S_x(t)), \quad N^B(t) \sim Bin(N_0, S_{x-y}(t)), \tag{3.13}$$

where $S_x(t)$ is the survival probability given by the deterministic PH-aging model (see Eq. (2.4)). In this case, $N^A(t)$ and $N^B(t)$ are independent random variables. In order to introduce a correlation we may connect the number of survivors as follows

$$N^B(t) = N^A(t)\sigma_t, \qquad \sigma_t > 1. \tag{3.14}$$

Here, σ_t is a constant for a given t and is greater than one as cohort B is supposed to be healthier than cohort A. The choice of σ_t can be different; here, we find it logical to assume $\sigma_t = E[N^B(t)]/E[N^A(t)]$. Here, σ_t can be equivalently rewritten as

$$\sigma_t = S_{x-y}(t)/S_x(t). \tag{3.15}$$

It is obvious that the expectations of the number of survivors are

$$E[N^A(t)] = N_0 S_x(t), \qquad E[N^B(t)] = N_0 S_{x-y}(t). \tag{3.16}$$

The variances are given by

$$Var[N^A(t)] = N_0 S_x(t)(1 - S_x(t)), \qquad Var[N^B(t)] = \sigma_t^2 Var[N^A(t)]. \tag{3.17}$$

It is easy to see that

$$Cov[N^A(t), N^B(t)] = \sigma_t Var[N^A(t)]. \tag{3.18}$$

Method IV. We use a *common shock model* to introduce the dependence between the cohorts. The common shock model is useful to describe the dependence of lives when an exogenous event arises that affects each of the lives. Such models are common and well explained, for example, in Bowers et al. [10]. Here, we apply the technique to the number of survivors as follows. We suppose that the individuals of cohort B can be decomposed into two groups: the first group has the same health properties as the individuals of cohort A, the second group is independent from the first and has a better health. Mathematically, we have

$$\begin{aligned} N^A(t) &= X(t) + f_A(Y(t)), \\ N^B(t) &= Z(t) + f_B(Y(t)), \end{aligned} \tag{3.19}$$

where the random variables $X(t)$, $Y(t)$ and $Z(t)$ are mutually independent, f_A and f_B are deterministic functions of $Y(t)$. $X(t)$, $f_A(Y(t))$ and $f_B(Y(t))$ represent the number of survivors in the group with the health properties of cohort A; $Z(t)$ – with the health properties of cohort B. For the simplicity of calculations we assume that $f_A(Y(t)) = f_B(Y(t)) = Y(t)$; for general f_A and f_B the computations below can be carried out in an analogous way.

Assume $Y(0) = k$. Furthermore, let us adopt the Poisson assumption for the number of survivors, which is also very common and well justified in Brillinger [12]. In the deterministic phase-type aging model we have: $X \sim Pois((N_0-k)S_x(t))$, $Y \sim Pois(kS_x(t))$, $Z \sim Pois((N_0-k)S_{x-y}(t))$. Thus,

$$\begin{aligned} N^A(t) &\sim Pois(N_0 S_x(t)), \\ N^B(t) &\sim Pois(kS_x(t) + (N_0 - k)S_{x-y}(t)). \end{aligned} \tag{3.20}$$

It is easy to see that the expectation of the number of survivors in cohort A is the same as in the previous method,

$$E[N^A(t)] = N_0 S_x(t), \tag{3.21}$$

and for cohort B it is given by

$$E[N^B(t)] = kS_x(t) + (N_0 - k)S_{x-y}(t), \tag{3.22}$$

which is smaller than in the method I for all t. Due to the Poisson assumption, we have

$$Var[N^A(t)] = E[N^A(t)], \qquad Var[N^B(t)] = E[N^B(t)]. \tag{3.23}$$

Due to the mutual independency of X, Y and Z in Eq. (3.19), we have

$$Cov[N^A(t), N^B(t)] = Var[Y] = kS_x(t). \tag{3.24}$$

We demonstrate the simulation of the number of survivors in Fig. 3.2. Here, parameter μ of the gamma process γ_t equals 0.5; this parameter can be interpreted as a longevity risk level (see Lin and Liu [41]). The dashed grey lines represent the number of survivors $N^A(t)$ and $N^B(t)$, simulated using method I. We have shown analytically (in Eq. (3.2), Eq. (3.10), Eq. (3.16), Eq. (3.21)) that methods I-II and III-IV give the same expected values of $N^A(t)$. Same conclusions hold for the expectation of $N^B(t)$: all the methods give same values, except method IV, which gives a smaller value for all t (see Eq. (3.22)). The circles represent the expected values of $N^A(t)$ and $N^B(t)$, computed with methods I and II. The solid lines represent the expected values of $N^A(t)$ and $N^B(t)$, computed with methods III and IV. The observation that the circles and the solid lines (except method IV for $N^B(t)$) are very close to each other is explained by a small difference between Λ and $\tilde{\Lambda}$ for $\mu = 0.5$. For greater values of μ this difference becomes more significant.

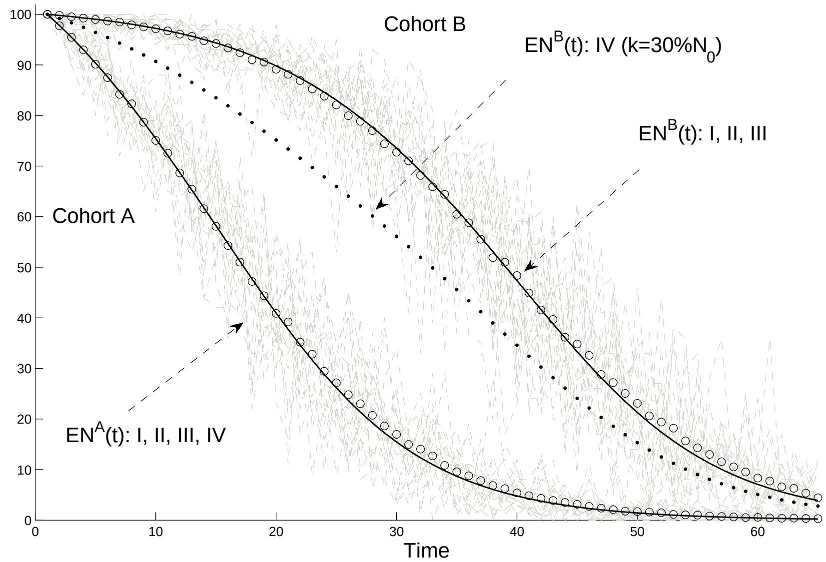

Figure 3.2: Number of survivors, methods I-IV
Parameters: $N_0 = 100, k = 0.3N_0, x = 65, y = 25$, SW1911M

3.3 Example: hedging with longevity bonds

The problem we consider in this section is described in Leppisaari [38] and serves as a model example of a situation, which requires computations of the number of survivors in two correlated cohorts. In order to introduce this problem it is useful to define the notion of longevity bonds.

Longevity bonds are financial instruments that help to hedge against longevity risk. Longevity bonds enable to use a small portion of current savings to buy guaranteed income for later years of retirement, providing financial security. Longevity bonds pay a coupon that is proportional to the number of survivors in a reference birth cohort (see [63]). Usually, the reference birth cohort is chosen on the basis of the lowest mortality rates. For example, often a Welsh cohort of males is considered as a reference cohort. We remark that the longevity bond is still an instrument under development and, despite a few attempts of its issuance, it has not reached the market so far (see Thomsen and Andersen [59]). However, we believe that the market will soon suggest a similar instrument against longevity risk, alternative to the longevity bond (see, for example, Barrieu et al. [8]).

Assume that the manager of a pension fund wishes to buy some longevity bonds to hedge the fixed population of annuitants of one unit lifetime pension against longevity risk. In this case, the future profit of the manager depends on the number of survivors in the two cohorts: cohort A, which consists of the annuitants, and cohort B, which is the "reference" cohort of the longevity bonds. The question is how many longevity bonds the manager has to buy. In [38], the author suggests to determine the optimal value of longevity bonds such that it minimizes the variance of future profits and losses.

The manager adopts the following investment strategy at time 0: to take a fixed-interest loan to buy longevity bonds with value N. Then, the total profit and loss of the manager at time t is given by

$$F(t) = (N^A(t) - E[N^A(t)]) - N(\delta N^B(t)/N_0 - r), \tag{3.25}$$

where r is a fixed interest rate, δ is % paid by the longevity bonds. The first term of the equation is the profit and loss brought by the annuitants themselves; the second term is the coupon paid by the longevity bonds minus the payment of interest to the bank. The variance of $F(t)$ is given by

$$Var[F(t)] = Var[N^A(t)] + \left(\frac{N\delta}{N_0}\right)^2 Var[N^B(t)] - 2\frac{N\delta}{N_0}Cov[N^A(t), N^B(t)]. \tag{3.26}$$

The value of N that minimizes $Var[F(t)]$ is, obviously, a function of t, and is given by

$$N^*(t) = \frac{N_0 Cov[N^A(t), N^B(t)]}{\delta Var[N^B(t)]}. \tag{3.27}$$

It is easy to see that if there is a perfect correlation between cohort A and cohort B, then $N^* = N_0/\delta$; if there is no correlation at all, then $N^* = 0$.

Using Eq. (3.6) and Eq. (3.8), (3.11) and (3.12), (3.17) and (3.18), (3.23) and (3.24), we obtain the values of $N^*(t)$ for the four methods, described in Section 3.2:

- **Method I**:

$$N^*(t) = \frac{N_0(\underline{\tau}_x \otimes \underline{\tau}_{x-y})B_t}{\delta((\underline{\tau}_{x-y} e^{\tilde{\Lambda} t}\mathbf{1} - (\underline{\tau}_{x-y} \otimes \underline{\tau}_{x-y})e^{\widetilde{(\Lambda \oplus \Lambda)} t}(\mathbf{1} \otimes \mathbf{1}))/N_0 + (\underline{\tau}_{x-y} \otimes \underline{\tau}_{x-y})B_t)},$$

or, equivalently,

$$N^*(t) = \frac{N_0 Cov[S_x^\gamma(t), S_{x-y}^\gamma(t)]}{\delta((E[S_{x-y}^\gamma] - E[S_{x-y}^\gamma]^2)/N_0 + Var[S_{x-y}^\gamma(t)])}; \tag{3.28}$$

- **Method II**:
$$N^*(t) = \frac{N_0(\underline{\tau}_x \otimes \underline{\tau}_{x-y})B_t}{\delta(\underline{\tau}_{x-y} \otimes \underline{\tau}_{x-y})B_t}, \tag{3.29}$$
or, equivalently,
$$N^*(t) = \frac{N_0 Cov[S_x^\gamma(t), S_{x-y}^\gamma(t)]}{\delta Var[S_{x-y}^\gamma(t)]}; \tag{3.30}$$

- **Method III**:
$$N^*(t) = \frac{N_0}{\delta \sigma_t},$$
or, using Eq. (3.15),
$$N^*(t) = \frac{N_0 S_x(t)}{\delta S_{x-y}(t)}; \tag{3.31}$$

- **Method IV**:
$$N^*(t) = \frac{N_0 k S_x(t)}{\delta(k S_x(t) + (N_0 - k) S_{x-y}(t))},$$
or, equivalently,
$$N^*(t) = \frac{N_0}{\delta \left(1 + \dfrac{N_0 - k}{k} \dfrac{S_{x-y}(t)}{S_x(t)}\right)}. \tag{3.32}$$

Let us consider a numerical example. In order to simplify the computation procedure, instead of the full matrix Λ that normally contains around 200 states, we take the artificially generated matrix $\hat{\Lambda}$ with five states given by

$$\hat{\Lambda} = \frac{1}{20} \begin{pmatrix} -0.4 & 0.2 & 0 & 0 & 0 \\ 0 & -0.5 & 0.4 & 0 & 0 \\ 0 & 0 & -0.6 & 0.3 & 0 \\ 0 & 0 & 0 & -0.7 & 0.4 \\ 0 & 0 & 0 & 0 & -0.9 \end{pmatrix} \tag{3.33}$$

One may verify using Eq. (1.28) that the average life expectancy at birth in such model equals 91.58 years. In Fig. 3.3 we demonstrate the variance of $F(t)$, computed using methods I-IV at $t = 10$, with respect to the value N. We observe that whereas methods II and III allow us to reduce the variance to almost zero, the variance in methods I and IV remains greater than a positive constant. We also observe that the methods II and III give us a similar value of the optimal nominal. As one can

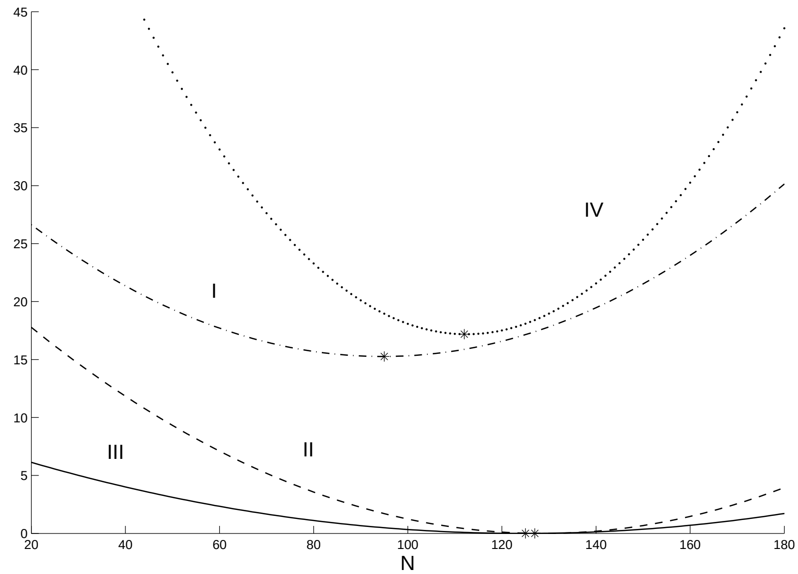

Figure 3.3: $Var[F(t)]$ vs N, $t = 10$, methods I-IV.
Parameters: $N_0 = 100,\ k = 0.9N_0,\ x = 65,\ y = 5,\ \mu = 3,\ \delta = 0.8,\ \Lambda = \hat{\Lambda}$.

see, under the considered parameters all methods I-IV give N^* which is greater or around N_0.

Another interesting observation is that the variance increases for a certain period of time, then decreases. Indeed, the increase in the first years is mostly caused by the uncertainty in the future mortality rates; in the long run there will be less and less individuals alive, which decreases this uncertainty. In Fig. 3.4 we present the values of the variance, evaluated at the optimal points, $N^*(t)$. In this figure one can also see that the variance is zero at time zero, which is expected, as we have no uncertainty at time zero. In Fig. 3.5 we present the variances in method II as a function of N, computed for different values of t.

In addition, in Fig. 3.5 we observe that $N^*(t)$ is not very sensitive with respect to t, which means that it is sufficient to solve the optimization problem only at one point of time.

It is a known fact that the increase of the number of individuals in a population diversifies the risk related to the uncertainty about the mortality of one individual. The effect of N_0 is seen by a comparison between Fig. 3.3 and Fig. 3.6, constructed with the same parameters, except that N_0 changes its value from 100 to 10000. We

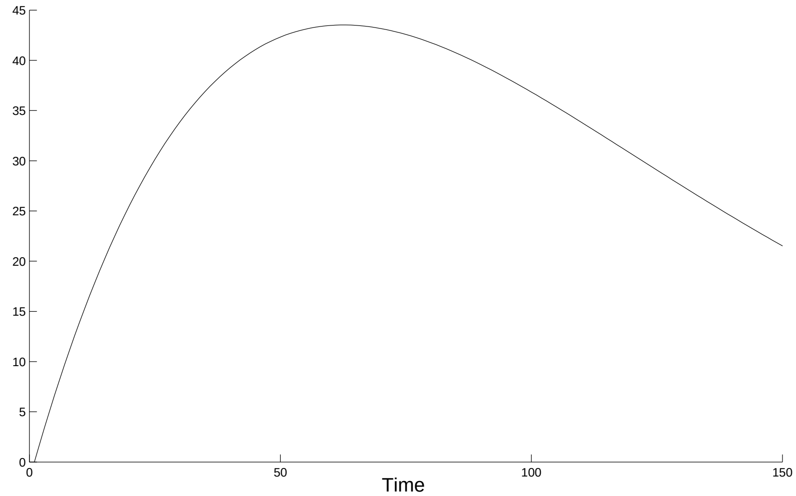

Figure 3.4: $Var[F(t)]$ at $N^*(t)$ vs t, method II.
Parameters: $N_0 = 100,\ k = 0.9N_0,\ x = 65,\ y = 5,\ \mu = 3,\ \delta = 0.8,\ \Lambda = \hat{\Lambda}$.

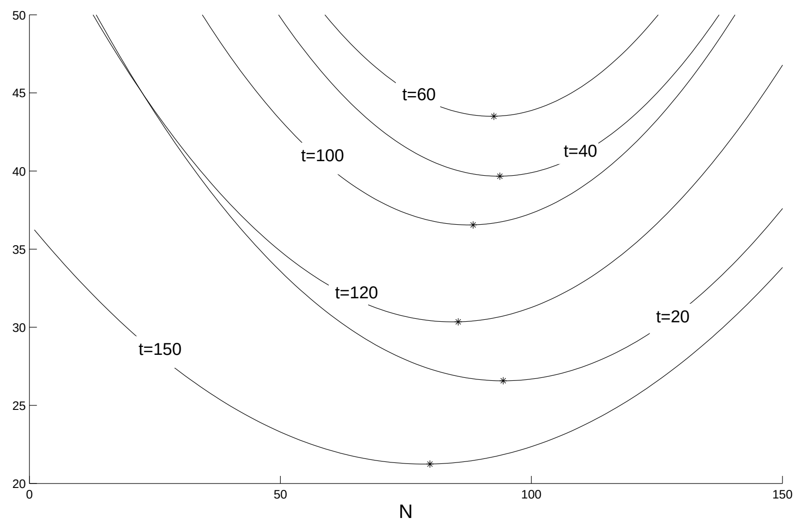

Figure 3.5: $Var[F(t)]$ vs N, method II.
Parameters: $N_0 = 100,\ k = 0.9N_0,\ x = 65,\ y = 5,\ \mu = 3,\ \delta = 0.8,\ \Lambda = \hat{\Lambda}$.

have two interesting observations: 1) the variance for methods I and II increases more, than for the methods III and IV; 2) the optimal value of N becomes visibly the same for methods I and II.

The first observation immediate follows the equations, obtained for the variances of $N^A(t)$ and $N^B(t)$ and their covariance. In methods III and IV, these quantities depend linearly on N_0, as shown in Eq. (3.17), (3.23) and Eq. (3.18), (3.24); in methods I and II the dependence on N_0 is quadratic, as shown in Eq. (3.6),(3.11), and Eq. (3.8),(3.12). This means that, for large N_0, $Var[F(t)]$, given by Eq. (3.26), has a linear dependence on N_0 in methods III and IV, and a quadratic dependence on N_0 in methods I-II.

The second observation is explained by Eq. (3.28) and Eq. (3.30), which show that $N^*(t)$ in method I and method II differs only by a constant, independent on N_0. Thus, for big values of N_0 the difference might not be noticeable in the figure.

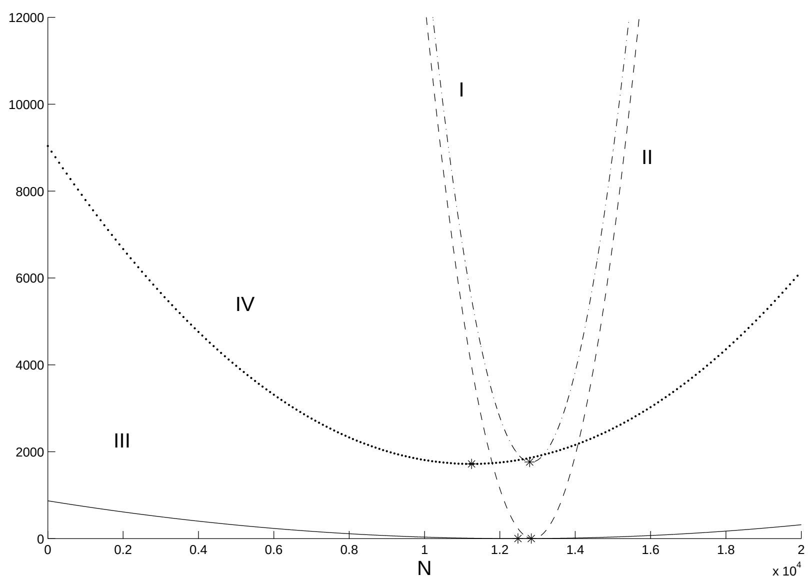

Figure 3.6: $Var[F(t)]$ vs N, methods I-IV.
Parameters: $N_0 = 10000,\ k = 0.9N_0,\ x = 65,\ y = 5,\ \mu = 3,\ \delta = 0.8,\ \Lambda = \hat{\Lambda}$.

Part II

Pension funds modeling

Summary of Part II

In this part, we employ the phase-type assumption for lifetime to examine the profitability of a pension fund.

Pension Funds can be described as life insurance companies selling two basic types of pension schemes: defined benefit and defined contribution. With *defined benefit* schemes (abbreviated as "DB") the pension and other benefits are set out in the contract. This implies the calculation of a premium for a defined pension and benefits. With *defined contribution* schemes (abbreviated as "DC") the contributions are fixed (often as percentages of the salary) and the benefits have to be determined.

Profit-test models have been recognized as a major tool available to actuaries involved in product development and risk management. A profit test uses projection mathematics to establish the prospective profit profile of a policy on a given set of assumptions. The *profit profile*, which is derived from a profit test, is the stream of profits, which flow from the policy over its lifetime (Bertschi et al. [9]). The development of profit profiles is well documented in papers such as Smart [56]. The resulting profit profile is discounted at a risk rate to give the *present value* of future profits. Formally, the *present value* of a cash flow at time t is a discounted difference between cash inflows, F_t^+, and outflows, F_t^-:

$$PV_t = (F_t^+ - F_t^-)v^t, \tag{3.34}$$

where v is a discount factor defined in Eq. (1.4) and Eq. (1.5). Recent profit-test models describe two types of pension schemes: *open* and *closed*. In open pension schemes, there are new plan participants who arrive to the pension plan according to a chosen law. Closed pension schemes describe one fixed population of plan participants.

Not only do profit-test models play a key role in the development of new pension schemes and estimation of their profitability, but they also serve as a tool to improve a pension scheme. Also, this type of models is useful for the purposes of risk management and implementation of new regulations.

In Bertschi et al. [9], Janssen and Manca [34] and Mettler [44], the authors construct profit-test models to estimate the expected present value of future cash flows of a pension fund. In these papers, pension plan participants are described by a Markov chain with a subjectively chosen state space (we give a more detailed description in Section 1.4). Pension plan contribution rates and benefits are considered as exogenous parameters of the profit-test model, independently whether the plan is of DB or DC type. Cash inflows consist of contributions to active plan participants and cash outflows consist of benefit payments and expenses. For example, in [44] there are benefits to disabled and retired beneficiaries, lump sum payments to widows and orphans as well as the payment of vested benefits to resigned beneficiaries. Salaries of plan participants in these papers are also parameters of the models and often linearly increase with time. In [9], [34] and [44] both open and closed pension schemes are considered.

In Chapter 4 we construct a profit-test model for a DB pension plan. Similarly to [9], [34] and [44] we employ a Markov chain to model pension plan participants. The new assumption that we make is that when active, the participants are described by a number of health states and evolve from one health state to another in accordance with the PH-aging model. We have three categories for non-active participants: retired, resigned (termed "surrendered" in the sequel) and dead. As we have discussed earlier, the phase-type lifetime assumption allows us to incorporate health as a factor and characterize the participants in a more precise manner. However, we can not apply the assumption straightaway, and we need to modify the PH-aging model to include additional causes of decrement such as retirement and surrender. In our model we make an assumption that the surrender rate is a deterministic function of seniority, which is often used in practice, and that pension plan participants behave independently from each other. In [43], the authors model surrenders stochastically, allowing for dependent behavior of policyholders.

We model an open pension scheme. Specifically, we assume that, when the participants leave the pension plan due to one or another reason, there are new plan participants with the same initial individual characteristics that come as their replacement. The replacement is assumed to happen after some delay, which is considered as a parameter of the model. Clearly, by setting the delay to be infinite, one arrives to a closed pension scheme. Therefore, our chosen pension plan exhibits some properties of an open pension scheme and at the same time can be easily reduced to a closed one.

As we have chosen a DB pension plan we need to specify the benefits and determine the contribution rates. Firstly, we assume that plan participants receive a lifetime pension upon retirement. We remark that in Chapter 4 we do not follow plan participants after retirement: we assume that at the moment of retirement they receive a lifetime annuity as a lump sum and disappear from our consideration. Secondly, in case of death before the retirement age, or if a participant leaves the plan for personal reasons, the pension fund reimburses the sum of all accrued premiums. The assumption that the benefits in case of death or surrender are equal to the sum of accrued premiums is motivated by the following reason: in order to determine the amount of the benefit for a randomly taken participant at time t, one needs to determine the number of years that this participant has spent in the pension plan (termed "seniority" in the sequel) and the contribution rate for this participant, which in our model depends on his/her health state at the time of arrival. Our purpose is to demonstrate that the phase-type approach allows us to keep this information and, therefore, allows us to deal with benefits of such complex structure. Unlike in [9], [34] and [44], the contribution rates in our profit-test model are not exogenous parameters, we obtain equations to compute them for each age and each health state.

Similarly to the existing studies, cash inflows are contributions from active plan participants, and cash outflows are benefit payments and expenses. Other technical assumptions are that the initial population of plan participants is uniformly distributed between the ages x_l and x_u, and that the fund has both recurrent, and lump-sum expenses per policy. These assumptions do not affect the principles of the profit-test model construction. At the end of Chapter 4, in a similar manner as in Bertschi et al. [9], we examine the effect of different replacement policies of the fund by varying the speed at which non-active plan participants are being replaced.

In Chapter 5, we perform a risk analysis of the cash flows obtained in Chapter 4. Profits may decrease due to causes of very different nature. Obviously, the fact of increasing life expectancy (Oeppen and Vaupel [47]) leads to a longer period of pension payments and lowers the profits. Another potential source of losses is due to unpredicted changes on the market, for example, a fall of investment rates. Thus, apart from the basic risk related to current mortality, we find it important to consider longevity and market risks. For the purpose of longevity risk evaluation, we obtain the distribution of the present values and perform a perturbation analysis with respect to future mortality rates. In order to evaluate the market risk, we consider two new definitions of the cash flows: one with investment benefits (abbreviated as "IB") on the accumulated contributions, and the other one with IB on both the accumulated contributions and the previous IB. In all these cases we model the investment rate as a *Markov reward process*, as suggested in Norberg [46].

At the end of Chapter 5, we are interested in analysis of the pension system in the long-term. We determine the time to stability, maximal and average seniority in the plan and the stationary health state distribution of plan participants. We analyze the impact of the change of mortality rates on the stationary distribution.

We find it natural to assume that the cash flows in the post-retirement period significantly depend on the health and, therefore, on the future mortality rates of plan participants. For this reason, in Chapter 6 we consider the pre-retirement period as static, and we focus on the impact of health on the distribution of the future cash flows coming from the pensioners. Here, similarly to Chapter 4, we construct a profit-test model for an open pension scheme, and we assume that new participants are participants at retirement who arrive to the fund over the years. The pensioners are assumed to have the same age R at the time of their arrival, and each brings to the fund the accumulation $\ddot{a}_R$ corresponding to a lifetime pension of size 1. The first pension payment is made at the time of arrival. We assume that the pensioners evolve in time in accordance with the PH-aging model, and can no longer retire (obviously, as they are pensioners already) or surrender.

The present value of the future cash flows is defined as the discounted sum of all lifetime annuities transferred to the fund by the pensioners at the time of their retirement, minus discounted pension payments made by the fund to survivors. Furthermore, we suppose that when a pensioner dies his remaining funds are left in the fund to finance pension obligations with respect to other pensioners. The remaining funds are negative if a pensioner lives longer than expected, that is, if the total payments to the pensioner exceed the lifetime annuity. It must be emphasized that our specific definition of the cash flows is not the main focus and one might choose other definitions; the phase-type approach is the main point.

To obtain the distribution of the present value of the cash flows, we determine the distribution of an individual contribution, and we consider two different models of the arrival of new pensioners. In the first model, we assume a constant number of new pensioners joining the fund each year; in the second model, we assume that new pensioners arrive according to a Poisson process. For the first model, we approximate the distribution of the total present value by a normal distribution. In the second model we follow a more detailed procedure, because in this case, the total present value is a compound Poisson random variable. One of the traditional ways to obtain the distribution of such a random variable is to apply the recursive algorithm introduced by Panjer [49]. As we have discussed above, the individual present value may be negative, and therefore we modify the algorithm on the basis of the extension of Panjer's algorithm given in Sundt and Jewell [58]. We discuss the efficiency of this approach by performing a comparison between the resulting distribution of total present value and its normal approximation. We evaluate the impact of health on the distribution for the models with deterministic and stochastic arrivals by applying the longevity risk estimation tools developed in Section 2.4. The main aspects of this work are presented in Govorun and Latouche [28].

As our main contribution in Part II, we indicate the development of the methodology to estimate the profitability of the suggested open DB pension plan, where the lifetime and health of plan participants are described by a phase-type distribution.

The estimation of the profitability focused on the pre-retirement period has required the elaboration of a profit-test model. To construct the profit-test model we have developed methods to extend the PH-aging model to multiple decrements such as retirement and surrender, and to determine contribution rates for ages and health states. In order to correctly estimate future cash flows, we have determined the distribution of the seniority of plan participants and the distribution of their entrance health state. This allowed us to determine the distribution of the present value of the future cash flows. In addition, we have proposed various techniques for longevity and market risk estimations, and a technique to perform the stability analysis of the pension plan. We have also suggested a numerical procedure to estimate the financial impact of the replacement/recruitment policy. Here, we obtained the optimal speed at which the leaving plan participants have to be replaced in order to balance cash inflows and outflows in the long term.

To estimate the profitability of the post-retirement period we have elaborated the technique for the construction of two profit-test models, with deterministic and with stochastic arrivals. The comparative analysis of these two models has justified the use of the model with stochastic arrivals over the model with deterministic arrivals on a short-term horizon. We have determined the distribution of the present value of future cash flows and closed form expressions for the moments. This enabled us to develop a technique to analyze the financial impact of health on the distribution of the present value of future cash flows. The elaborated technique has allowed us to compare the financial impact of two different events: early retirement and the reduction of future mortality rates.

Chapter 4

Profits&Losses: pre-retirement

We begin this chapter by Section 4.1, where we configure the mathematical model for the pension plan participants. Specifically, we show how to extend the PH-aging model to allow for other decrements. In Section 4.2 we use the phase-type approach to determine pension plan contribution rates, which we compare to the values computed with the classical approach. In Section 4.3 we obtain the profit profile of the pension plan and perform two simple verification procedures. In Section 4.4 we investigate the possibility to correct the pension plan to achieve better profitability by changing the arrival rate of new plan participants.

4.1 Plan participants

We follow the idea described, for example, in Bertschi et al. [9] and in Mettler [44] to model pension plan participants using a Markov chain. We assume that the Markov chain has $n + 3$ phases and that the states are grouped into four categories:

- States 1 to n states describe *active participants* (A);
- State $n + 1$ corresponds to *retired participants* (R);
- State $n + 2$ corresponds to *surrendered/resigned participants* (S);
- State $n + 3$ corresponds to *deceased participants* (D).

We look at the evolution of plan participants in continuous time and organize the structure of the generator matrix according to these four sets of states. We assume the generator matrix Π to have the form:

$$\Pi = \begin{pmatrix} \Pi_{AA} & \underline{q}^{\mathrm{T}}_{\{R\}} & \underline{q}^{\mathrm{T}}_{\{S\}} & \underline{q}^{\mathrm{T}} \\ \underline{d} & -\phi & 0 & 0 \\ \underline{d} & 0 & -\phi & 0 \\ \underline{d} & 0 & 0 & -\phi \end{pmatrix} \tag{4.1}$$

where Π_{AA} is an n by n matrix that describes active participants,

$$\Pi_{AA} = \Lambda - diag\left(\underline{q}^{\mathrm{T}}_{\{R\}}\right) - diag\left(\underline{q}^{\mathrm{T}}_{\{S\}}\right), \tag{4.2}$$

where Λ is the transition rate matrix (2.2) for the PH-aging model. $\underline{q}^{\mathrm{T}}_{\{R\}}, \underline{q}^{\mathrm{T}}_{\{S\}}$ and $\underline{q}^{\mathrm{T}}$ are column vectors of size n representing transitions to one of the inactive states ($\underline{q}^{\mathrm{T}}$ is given in Eq. (2.1)), ϕ is a parameter that represents the rate at which a plan participant who has left the pension plan is being replaced. $\underline{d}$ is the row vector of size n describing replacements in case of retirement, surrender and death. The diagrammatic representation of the proposed model is presented in Fig. 4.1.

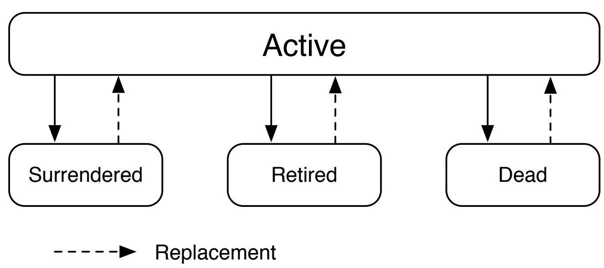

Figure 4.1: Modeling of pension plan participants

Denote the state at time t as Φ_t and the number of years spent in the system at time t as an active participant as Ψ_t. We calculate all financial results once in a year until a chosen time horizon, implying a discrete evolution of plan participants. In accordance with Eq. (1.24) the state distribution at time $t+1$ is given by

$$\underline{p}_{t+1} = \underline{p}_t e^{\Pi}, \quad t \geq 0,$$

where $\underline{p}_t = (p_t^{(i)} : i = 1, \ldots, n+3)$, $p_t^{(i)} = \mathrm{P}\left[\Phi_t = i\right]$. We assume that matrix Π is irreducible and finite, therefore, there is always stationary distribution given by $\lim_{t \to \infty} \underline{p}_t$.

The probability that a new participant enters the pension plan at time t in state i is defined by

$$M_t^{(i)} = \mathrm{P}\left[\,\Psi_t = 0\,,\ \Phi_t = i\,\right] \tag{4.3}$$

where $i \in A$. Define $Q = \{R, S, D\}$, then we can determine $M_t^{(i)}$ as

$$M_t^{(i)} = \underline{p}_{t-1}^{(Q)}\left(e^{\Pi}\right)_{Qi}, \tag{4.4}$$

where

$$\underline{p}_t^{(Q)} = \left(p_t^{(i)},\ i \in \{R, S, D\}\right) \tag{4.5}$$

Let us denote the distribution of states for active participants as $\underline{p}_t^{(A)}$. It is given by

$$\underline{p}_t^{(A)} = \left(p_t^{(i)},\ i \in \{A\}\right) \tag{4.6}$$

The expected number of active plan participants at time t is equal to $N\underline{p}_t^{(A)}\mathbf{1}$, where N is the total number of plan participants defined from the start of the plan. This number converges to a constant due to the convergence of the distribution $\underline{p}_t^{(A)}$.

In order to complete the description of the evolution of plan participants we need to specify the transition rates of the matrix Π and the initial distribution of states.

Initial distribution of health states. As was mentioned in the introduction to this chapter, we consider a population with uniformly distributed ages between x_l and x_u. Let us also assume that the population is of male gender and that they follow SW1911M. What is the corresponding state distribution $\underline{p}_0$ of this population? Recall that $\underline{p}_0$ is a vector of size $n+3$ and it can be expressed by

$$\underline{p}_0 = \left(\underline{p}_0^{(A)}, 0, 0, 0\right), \tag{4.7}$$

because there is no inactive participants at the launch of the plan. Denote the random value representing a real age at time zero by B and recall the vector $\underline{\tau}_x$ introduced in Eq. (2.3)

$$\underline{\tau}_x = \frac{\underline{\alpha} e^{\Lambda x}}{\underline{\alpha} e^{\Lambda x}\mathbf{1}}, \tag{4.8}$$

which gives the distribution of Φ_0, the phase process at time $t = 0$ given the age is x, $(\underline{\tau}_x)_i = \mathrm{P}\left[\,\Phi_0 = i\,|\,B = x\,\right]$, $i \in A$. In order to obtain $\underline{p}_0^{(A)} = \left(p_0^{(j)},\ j \in A\right)$, we take into account the uniform distribution of initial ages. Thus, the initial phase distribution is given by

$$p_0^{(j)} = \mathrm{P}\left[\,\Phi_0 = j\,\right] = \frac{1}{x_u - x_l}\sum_{x=x_l}^{x_u}\left(\underline{\tau}_x\right)_j, \quad j \in A \tag{4.9}$$

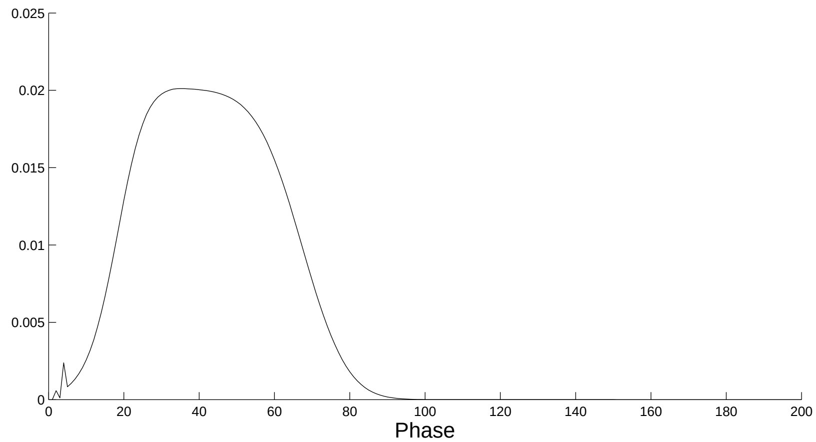

Figure 4.2: Health state distribution of active plan participants, $\underline{p}_0^{(A)}$
$x_l = 10$, $x_u = 30$, SW1911M

We illustrate the distribution $\underline{p}_0^{(A)}$ in Fig. 4.2.

The probability distribution of the real age given a health state is given by

$$\mathrm{P}\left[\,B = x \,|\, \Phi_0 = j\,\right] = \frac{\mathrm{P}\left[\,\Phi_0 = j \,|\, B = x\,\right]\mathrm{P}\left[\,B = x\,\right]}{\mathrm{P}\left[\,\Phi_0 = j\,\right]}$$

which leads to the expression

$$\mathrm{P}\left[\,B = x \,|\, \Phi_0 = j\,\right] = \begin{cases} \left(\underline{\tau}_x\right)_j \;\Big/ \sum_{t=x_l}^{x_u} \left(\underline{\tau}_t\right)_j, & x \in [x_l, x_u] \\ 0, & x \notin [x_l, x_u] \end{cases}, \qquad (4.10)$$

if B is uniform in $[x_l, x_u]$. We illustrate the probability $\mathrm{P}\left[\,B = x \,|\, \Phi_0 = j\,\right]$ for different j in Fig. 4.3. Here, as well as in all our numerical examples, we suppose that $x_l = 10$ and $x_u = 30$ as this allows us to have a longer follow-up period before the individuals retire. One may see from Fig. 4.3 that the probability mass indeed lays between x_l and x_u. Consider an individual in health state i for small i. Logically, the probability that the individual has the minimum possible age is rather high. The same situation holds for individuals who have high health states. This explains the peaks which one can see at the boundaries. The other real age distributions correspond to intermediate health states.

Transition rates: aging and death. We use the PH-aging model introduced in section 2.1 to obtain the aging rates and the death rates for active plan participants. Thus, the PH-aging models gives us Λ in Eq. (4.2), and vector $\underline{q}$ of mortality rates.

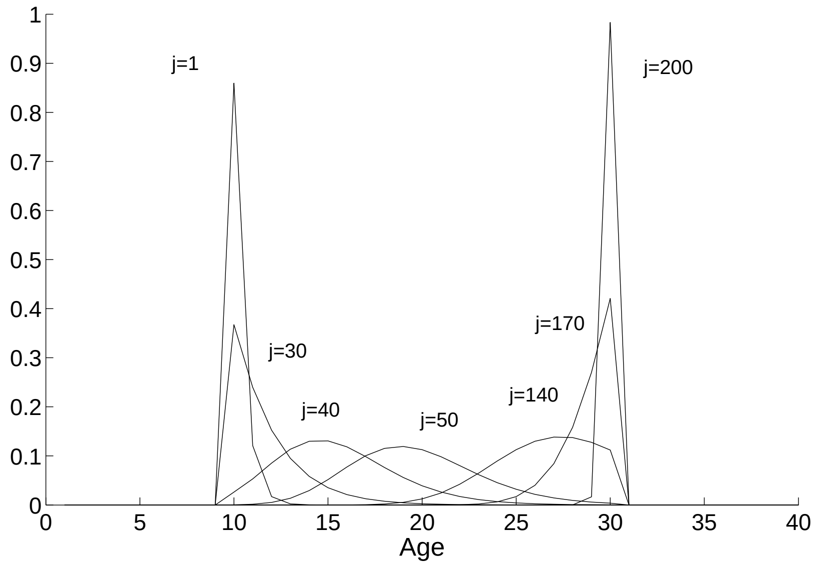

Figure 4.3: Conditional age distribution, $\mathrm{P}\,[\,B = x \,|\, \Phi_0 = j\,]$
$x_l = 10$, $x_u = 30$, SW1911M

According to Eq. (1.26), the probability to stay alive at least t years for a newborn individual (survival probability) is defined by

$$S^A(t) = \underline{\alpha} e^{\Lambda t} \mathbf{1}.$$

The survival function $S^A(t)$ has already been given in Fig. 2.3 and will be discussed again in Fig. 4.5 (crosses).

Transition rates: retirement. We assume that the retirement happens at a statutory retirement age, R. In order to determine the retirement rates $\underline{q}_{\{R\}}$ we have defined a model similar to the PH-aging model, but with an additional absorbing state representing retirement. The transition rate matrix of the new model is given by

$$\Lambda^R = \Lambda - diag\left(\underline{q}^{\mathrm{T}}_{\{R\}}\right). \tag{4.11}$$

The probability to stay alive and active for at least t years for a newborn individual (survival probability) is denoted as $S^{AR}(t)$, it is given by

$$S^{AR}(t) = \underline{\alpha} e^{\Lambda^R t} \mathbf{1}.$$

Define the jump function $F = F(t)$ such that

$$\begin{cases} F(t) = S^A(t), & t \in [0, R) \\ F(t) = 0, & t \geq R \end{cases}$$

We define $\underline{q}_{\{R\}}$ so that $S^{AR}(t)$ is an approximation of the function $F(t)$.

Denote by i^* the expected health state at age R. It is determined by

$$i^* = \sum_{i=1}^{n} i(\underline{\tau}_R)_i. \tag{4.12}$$

We assume the vector $\underline{q}_{\{R\}}$ to have the following structure:

$$\left(\underline{q}_{\{R\}}\right)_j = \begin{cases} r_1, & j \leq i^* \\ r_1 + r_2, & \text{otherwise}, \end{cases},$$

$j \in A$, and we apply the least squares method as a fitting procedure to determine r_1 and r_2. A rather good approximation is achieved when $r_1 = 0$ and r_2 is very high. The survival function $S^{AR}(t)$ is presented in Fig. 4.5 in small dots. The vertical line represents the jump of $F(t)$ at $R = 65$.

Transition rates: surrenders. Surrender rates $\underline{q}_{\{S\}}$ for active employees should be chosen to match actual data. We assume that one can obtain the empirical distribution of the number of years spent in the pension plan at the moment of stopping the contract. In our experience, the form of this distribution is quite stable for many pension plans. Therefore, we can derive an empirical survival probability, $\hat{S}_t$, which is the fraction of the plan participants who stay more than t years as active participant. This information has to be translated into the health state model. We give an example of $\hat{S}_t$ in Fig. 4.4 with the maximal number of years of service being approximately 12 years.

In order to find a surrender rate for every health state we define a model similar to the retirement model described above, but with an additional absorbing state representing surrenders. Let us denote the new phase-type distribution as $(\underline{\alpha}, \Lambda^S)$, where

$$\Lambda^S = \Lambda - diag\left(\underline{q}_{\{R\}}^{\mathrm{T}}\right) - diag\left(\underline{q}_{\{S\}}^{\mathrm{T}}\right).$$

We see from Eq. (4.2) that $\Lambda^S = \Pi_{AA}$. Recall that the initial distribution of the health states of plan participants is given by $\underline{p}_0$ and that $\left(\underline{p}_0 e^{\Lambda^S t}\right)_j$ is the probability to remain at least t units of time in the pension plan as an active participant and to be in state j at time t. We determine $\underline{q}_{\{S\}}$ so that $\underline{p}_0 e^{\Lambda^S t}\mathbf{1}$ is an approximation

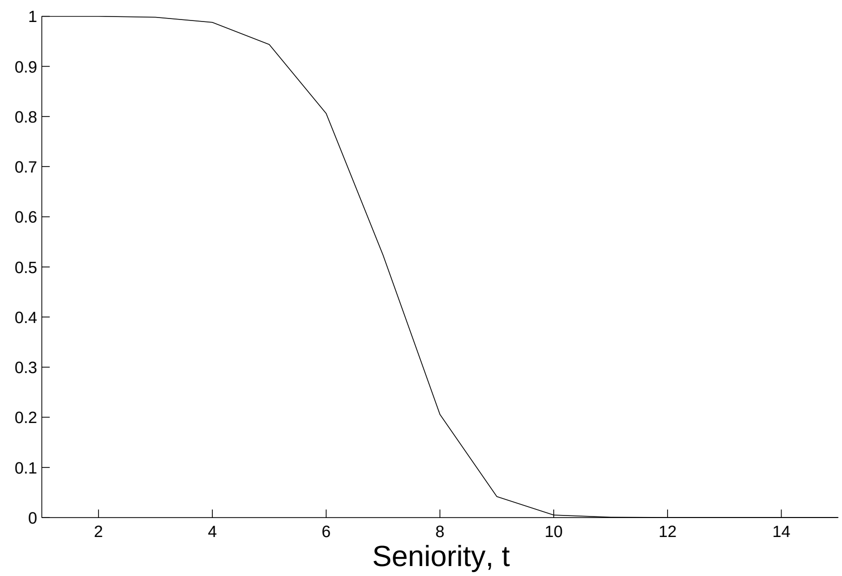

Figure 4.4: Example of empirical survival probability $\hat{S}_t$, $t = $ 1-15

of the empirical probability $\hat{S}(t)$. Like in the problem of retirement rates, we select a parametrized structure of vector $\underline{q}_{\{S\}}$ and apply the least squares method to fit $\underline{p}_0 e^{\Lambda^S t}\mathbf{1}$ to $\hat{S}(t)$ for all t.

Specifically, we assume vector $\underline{q}_{\{S\}}$ to hold a similar structure as for retirement rates. However, there are some differences. Denote the maximum possible length of service in the pension plan obtained from the data as l^* and let us define ϕ^* as the expected state of an individual who spent l^* years in the system with retirements, described above, at the time of retirement. The structure of vector $\underline{q}_{\{S\}}$ is defined by

$$\left(\underline{q}_{\{S\}}\right)_j = \begin{cases} w_1, & j \leq \phi^* \\ w_1 + w_2, & \text{otherwise} \end{cases}, \quad j \in A.$$

The main difference between this model and the previous two is the interpretation of time. In the model with surrenders, time has the meaning of the number of years spent in the pension plan and is no longer directly interpreted as the age of a participant. In order to compare this model to the previous two, we observe the survival functions for a newborn individual. For the first two models we can simply compare $S^A(t)$ and $S^{AR}(t)$, because at the start, all people are newborn and the time is the age of an individual. In case of the model with surrenders, time starts at the moment of entering the pension program, therefore, in order to perform a proper comparison, we need to take into account the age at which an individual enters the plan as an active participant. Let this age be a random variable B and let us pretend that every individual in the population joins the plan, if he/she lives until age B. In this way, our comparison with $S^A(t)$ and $S^{AR}(t)$ will not be influenced by any assumption about the fraction of the population who joins the pension plan.

Define by $S^{ARS}(t)$ the probability for a newborn to be in an active state at age t in the model with surrenders. Here, we need to underline that the function $S^{ARS}(t)$ is obtained only for demonstration purposes and does not affect our further calculations.

We obtain $S^{ARS}(t)$ in a step by step manner. Obviously, for a newborn individual $S^{ARS}(0) = 1$. At age $t = 1$, the individual either is insured for one year with probability $\mathrm{P}\,[\,B = 0\,]$ or does not have any insurance with probability $\mathrm{P}\,[\,B > 0\,]$. This implies that

$$S^{ARS}(1) = \mathrm{P}\,[\,B = 0\,]\,\underline{\alpha} e^{\Lambda^S}\mathbf{1} + \mathrm{P}\,[\,B > 0\,]\,\underline{\alpha} e^{\Lambda}\mathbf{1}.$$

At age $t = 2$, there are three possible scenarios for the individual: not being a part of the pension plan with probability $\mathrm{P}\,[\,B > 1\,]$, being a plan participant for one year with probability $\mathrm{P}\,[\,B = 1\,]$ and being a plan participant for two years with probability $\mathrm{P}\,[\,B = 0\,]$. This results to

$$S^{ARS}(2) = \mathrm{P}\,[\,B = 0\,]\,\underline{\alpha} e^{\Lambda^S 2}\mathbf{1} + \mathrm{P}\,[\,B = 1\,]\,\underline{\alpha} e^{\Lambda} e^{\Lambda^S}\mathbf{1} + \mathrm{P}\,[\,B > 1\,]\,\underline{\alpha} e^{\Lambda 2}\mathbf{1}.$$

We follow the same argument to obtain the survival function at time t

$$S^{ARS}(t) = \sum_{x=0}^{t-1} \mathrm{P}\,[\,B = x\,]\,\underline{\alpha} e^{\Lambda x} e^{\Lambda^S (t-x)}\mathbf{1} + \underline{\alpha} e^{\Lambda t}\mathbf{1} \sum_{x=t}^{100} \mathrm{P}\,[\,B = x\,]\,. \tag{4.13}$$

The three survival functions $S^{ARS}(t)$, $S^A(t)$ and $S^{AR}(t)$ are presented in Fig. 4.5 in different types of dots. One may see from the figure, that $S^{ARS}(t)$ significantly differs from $S^A(t)$ and $S^{AR}(t)$. This is due to the fact that for this figure the initial distribution of real ages $\mathrm{P}\,[\,B = x\,]$ was chosen uniformly distributed between the ages 10 and 30 and the empirical survival probability $\hat{S}_t$ was constructed so that l^* is about 15 years. This implies that by $30 + 15 = 45$ years old the probability to survive in the system should be almost zero, which is observed from the figure. In Fig. 4.5, as well as in our further illustrations $\phi^* = 31$ and $w_1 = 0$, $w_2 = 0.1576$, unless the surrender option is specifically cancelled.

Transition rates: replacements. The last step is to define the last three lines of the generator matrix Π. The components have the interpretations of a physiological age structure of replacements and a replacement speed. As we have already

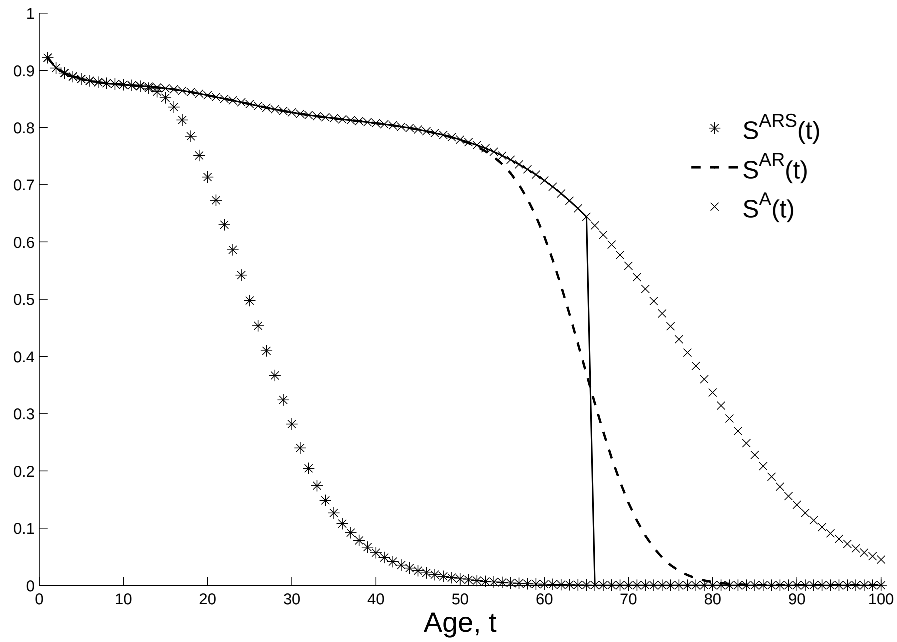

Figure 4.5: Survival probabilities, $S^A(t)$, $S^{AR}(t)$ and $S^{ARS}(t)$, SW1911M

assumed in Eq. (4.1), the speed of the replacements is the same for surrendered, retired and deceased participants and is determined by ϕ. We assume parameter ϕ, $\phi > 0$ to be an input parameter of the model. Clearly, the value of the parameter can be easily found from available data.

Furthermore, as we have discussed in the introduction to this chapter, we assume that the health structure at replacement is identical to the initial distribution of participants $\underline{p}_0$. Therefore,

$$\underline{d} = \underline{p}_0^{(A)}\phi.$$

4.2 Contribution rates

We use the traditional balance approach to calculate the contribution rates (we use below terms such as "tariffs" or "premiums") of the pension plan. We assume that an individual in state i at the moment t has the salary $\Theta_t^{(i)}$ and we calculate the tariff as a percentage of this salary. We also assume the pension fund to have two types of expenditures per policy: c, a recurrent cost per year and I_0, a one-time initial cost. In the balance approach we take into account two decrements: death and the exercise of the option to surrender. Obviously, the two decrements are dependent, which can be easily captured by the phase-type model.

Our model for the evolution of the plan participants uses health states instead of real ages. First of all, it makes the delay until the statutory retirement age random which implies an undefined horizon for calculations. Secondly, unlike for the mortality table survival probabilities, for an individual in state j, in the phase-type aging model the probability to survive for t years is not a product of successive one-year survival probabilities for the states from i to $(i+t-1)$. To deal with the difficulties we calculate gross premiums $\mu_x(i)$ for a contract which starts at age x in state i. The tariff $\mu_x(i)$, in our setup, is determined at the start of the contract and does not change for the whole duration of the contract.

Denote by ${}_{n,k}p_i^{[SD,D]} = \underline{\alpha}^{(i)} e^{\Lambda^{SD} n} e^{\Lambda k} \mathbf{1}$ the probability to remain active in the plan for n years with respect to death and surrender option, then remain alive for k years. The quantity ${}_{n}\left(p_q\right)_i^{[SD]} = \underline{\alpha}^{(i)} e^{\Lambda^{SD} n} (\mathbf{1} - e^{\Lambda^{SD}} \mathbf{1})$ represents the probability to remain active for n years, then become inactive due to death or surrender. Here, $\underline{\alpha}^{(i)} = (\alpha_j^{(i)}, j \in A)$ is introduced in Eq. (2.11) and Λ^{SD} is defined by removing retirement as a cause to leave the pension plan

$$\Lambda^{SD} = \Lambda - diag\left(\underline{q}^{\mathrm{T}}_{\{S\}}\right) = \Pi_{AA} + diag\left(\underline{q}^{\mathrm{T}}_{\{R\}}\right).$$

Define

$$
{}_{n|}\ddot{a}_i = \sum_{k=0}^{\infty} v^{n+k}\, {}_{n,k}p_i^{[SD,D]}, \quad \ddot{a}_{i:n}^{[\theta]} = \sum_{k=0}^{n} v^k\, {}_{k,0}p_i^{[SD,D]}\, \Theta_k^{(i)}, \tag{4.14}
$$

$$A_{i:n}^{[\theta]} = \sum_{k=1}^{n} v^k\, {}_{k-1}\left(p_q\right)_i^{[SD]} \sum_{j=0}^{k-1} \Theta_j^{(i)}.$$

With the introduced notations $\mu_x(i)\ddot{a}_{i:R-x}^{[\theta]}$ is the actuarial present value of the contributions accumulated over $R-x$ years by an individual aged x at the start of the pension plan, ${}_{R-x|}\ddot{a}_i^{[\theta]}$ is his/her lifetime annuity deferred by $R-x$ years, $\mu_x(i)A_{i:R-x}^{[\theta]}$ is the actuarial present value of the benefit in case of death or surrender (the benefit is equal to the sum of accumulated contributions). Then, the balance equation for $\mu_x(i)$ is

$$
\mu_x(i)\ddot{a}_{i:R-x}^{[\theta]} = {}_{R-x|}\ddot{a}_i + \mu_x(i)A_{i:R-x}^{[\theta]} + c\mu_x(i)\ddot{a}_{i:R-x}^{[\theta]} + I_0, \tag{4.15}
$$

where $c\mu_x(i)\ddot{a}_{i:R-x}^{[\theta]}$ and I_0 are recurrent and one-time expenditures, respectively.

The tariff $\mu(i)$ for state i is the weighted sum of tariffs $\mu_x(i)$ over all x multiplied by the probability to enter the plan in the age $B = x$, conditioned on the state being

i:

$$\mu(i) = \sum_x \mu_x(i) \mathrm{P}\left[B = x \mid \Phi_0 = i \right].$$

If B is uniform on $[x_l, x_u]$ it follows from Eq. (4.10) that the tariff for state i is determined by

$$\mu(i) = \sum_{x \in [x_l, x_u]} \mu_x(i) \left(\underline{\tau}_x\right)_i \left(\sum_{t=x_u}^{x_u} \left(\underline{\tau}_t\right)_i \right)^{-1}. \tag{4.16}$$

To obtain the tariff for age x, we weigh $\mu_x(i)$ with probabilities to be in state i conditioned on the age being x

$$\mu_x = \sum_{i \in A} \mu_x(i) \left(\underline{\tau}_x\right)_i.$$

From Eq. (2.3) it follows that the tariff for age x is

$$\mu_x = \left(\underline{\alpha} e^{\Lambda x} \mathbf{1}\right)^{-1} \sum_{i \in A} \mu_x(i) \left(\underline{\alpha} e^{\Lambda x}\right)_i. \tag{4.17}$$

We have also considered a simplified method to compute the tariffs. In the circumstances, where one has two separate sets of survival probabilities, one for death, one for surrender, it is a standard practice to make the approximation that the two causes of removal are independent. This is because available tables often give mortality and surrender rates separately, which forces actuaries to assume their independence. Obviously, it creates errors in the calculations. Here, we determine premium values using this simplified method and estimate the approximation errors.

In fact, to calculate tariffs with this method it is sufficient to use an analogue of the standard balance equation with two independent causes of decrements with Eq. (1.14). In this case, the probabilities in Eq. (4.15) are approximated by

$${}_{n,k}\hat{p}_i^{[SD,D]} = {}_n p_i^S \; {}_{n+k} p_i^D, \quad {}_n \left(\hat{p}_q\right)_i^{[SD]} = {}_n p_i^S \; {}_n p_i^D - {}_{n+1} p_i^S \; {}_{n+1} p_i^D$$

where

$${}_n p_i^S = \underline{\alpha}^{(i)} e^{\Lambda^{S|D} n} \mathbf{1}, \quad {}_n p_i^D = \underline{\alpha}^{(i)} e^{\Lambda n} \mathbf{1},$$

The matrix $\Lambda^{S|D}$ is obtained from the matrix Π_{AA} by excluding death and retirement as the reason of leaving the system so that

$$\Lambda^{S|D} = \Pi_{AA} + diag\left(\underline{q}^{\mathrm{T}}\right) + diag\left(\underline{q}_{\{R\}}^{\mathrm{T}}\right) = \Lambda^{SD} + diag\left(\underline{q}^{\mathrm{T}}\right).$$

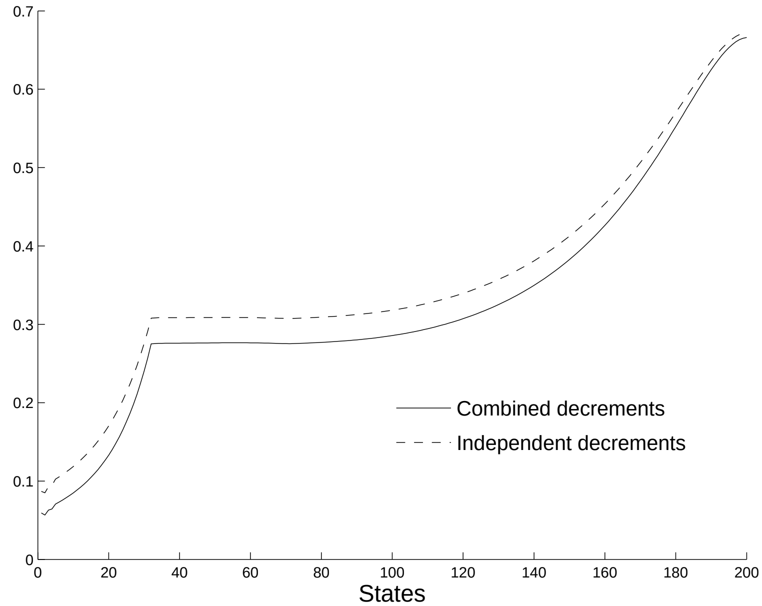

Figure 4.6: Tariffs $\mu(i)$ for health states
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.25$, SW1911M

We compare the two tariffs for the health states in Fig. 4.6. One may see from the figure that the difference between the two tariffs is noticeable, but is not very big. The change of the behavior of the curve corresponds to state 31: as we have mentioned in Section 4.1, in our examples the rate to surrender is zero for states below 31, otherwise it is equal to a positive constant. In Fig. 4.6, as well as in our further illustrations, salary $\Theta_t^{(i)} = 1$ for all t, i.

We compare the tariffs obtained for ages in Fig. 4.7. In addition to μ_x and $\hat{\mu}_x$ computed as indicated above, we also present the tariffs calculated from a mortality table, assuming independent decrements. Solid and dashed lines correspond to the phase-type model with dependent and independent decrements, respectively; the bullets correspond to classical mortality table approach.

4.3 Cash flows

Seniority distribution and reversal probability. To estimate the profitability of the pension plan, we need to estimate its future cash flows. Clearly, at every moment

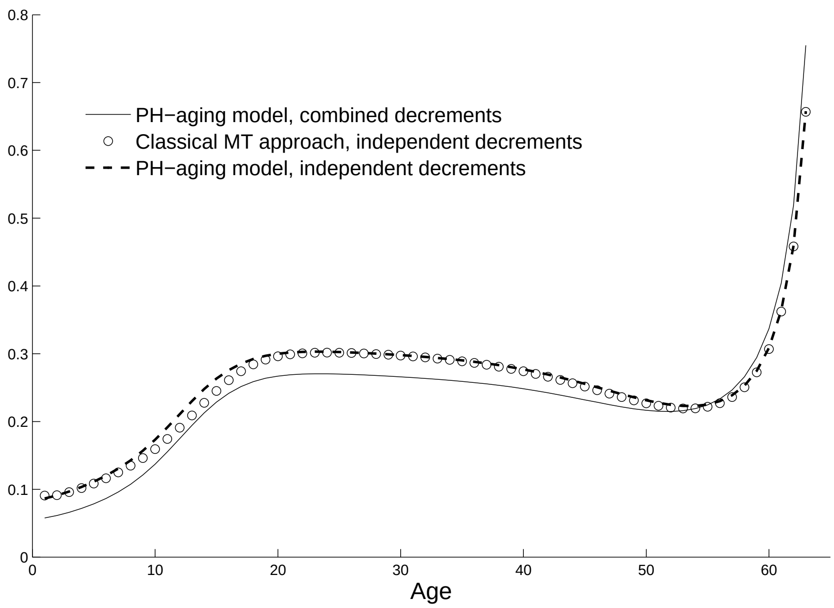

Figure 4.7: Tariffs μ_x for ages
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.25$, SW1911M

of time, the total cash flow is equal to the sum of the cash flows of all individual participants. To properly calculate the cash flow coming from one individual in state j at time t, we need to know how long the individual has already been in the system and what was his/her health state at the moment of entering the pension plan. Thus, for every individual we need the distribution of the number Ψ_t of years spent in the plan, that we call *seniority distribution*, and we need the distribution of the health state at the time of entrance, that we call *reversal probability*. Define the following event

$$\mathcal{E}_{ji}^{(r)}(t) = [\ \Psi_t = r\ ,\ \Phi_t = i\ ,\ \Phi_{t-r} = j\ ]. \tag{4.18}$$

$\mathcal{E}_{ji}^{(r)}(t)$ means that at time t a plan participant is in state i, $i \in A$ and he/she joined the pension plan at time $t-r$ being in state j, $j \in A$. Let us also define the joint probability of service r and state i at time t by ${}^{r}N_t^{(i)}$. It can be written as

$${}^{r}N_t^{(i)} = \mathrm{P}\,[\ \Psi_t = r\ ,\ \Phi_t = i\ ] = \sum_{j \in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right]. \tag{4.19}$$

Denote by ${}^r\mathbf{N}_t = \left({}^rN_t^{(i)},\ i \in A\right)$ the seniority distribution vector at time t. As suggested by Janssen and Manca [34], ${}^r\mathbf{N}_t$ can be computed from

$$\left\{\begin{array}{l} {}^r\mathbf{N}_{t+1} = ({}^{r-1}\mathbf{N}_t)\, e^{\Pi_{AA}} \\ {}^0\mathbf{N}_{t-r+1} = \mathbf{M}_{t-r+1} \end{array}\right. , \quad r \le t+1 \tag{4.20}$$

where $\mathbf{M}_{t-r+1}$ is the vector defined in Eq. (4.4), it is the vector of probability to be a new plan participant at time $(t-r+1)$. For the cash flow calculation we need conditional probabilities ${}^S_r\mathrm{P}_i(t) = \mathrm{P}\left[\,\Psi_t = r \mid \Phi_t = i\,\right]$ for active plan participants, which we find from the equation

$${}^S_r\mathrm{P}_i(t) \;\; = {}^rN_t^{(i)} \,/\, \left(\underline{p}_0^{(A)} e^{\Pi_{AA}t}\right)_i , \tag{4.21}$$

where $\left(\underline{p}_0^{(A)} e^{\Pi_{AA}t}\right)_i$ is the probability to remain among active plan participants for t consecutive years and to be in state i at time t. We demonstrate conditional seniority probabilities ${}^S_r\mathrm{P}_i(t)$ for different i in Fig. 4.8. In the figure, the time horizon t is equal to 100 and is greater than the stability time t^* that we compute later in Eq. (5.20). We observe from the figure that the higher the state, the higher the seniority. Also, despite in our examples the maximal observed seniority is $l^* = 15$, the distributions in the figure are concentrated on higher values of seniorities for high states. This is due to the fact that, as we can see from Eq. (4.21), the seniority probabilities are conditioned on the probability to be in active state i at time t.

The reversal probability to have entered the pension plan in state j, given that at time t the seniority is r and state is i, is ${}^R_r\mathrm{P}_{ji}(t)$ and is given by

$$\begin{aligned} {}^R_r\mathrm{P}_{ji}(t) \;\; &= \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right] / \textstyle\sum_{j\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right] \\ &= \mathrm{P}\left[\,\Phi_{t-r} = j\ ,\ \Psi_t = r\ ,\ \Phi_t = i\,\right] / \,{}^rN_t^{(i)} \\ &= \mathrm{P}\left[\,\Psi_{t-r} = 0\ ,\ \Phi_{t-r} = j\,\right] \left(e^{\Pi_{AA}r}\right)_{ji} / \,{}^rN_t^{(i)} \\ &= M_{t-r}^{(j)} \left(e^{\Pi_{AA}r}\right)_{ji} / \,{}^rN_t^{(i)}\ , \end{aligned} \tag{4.22}$$

where $\left(e^{\Pi_{AA}r}\right)_{ji}$ is the probability to stay among active participants for r consecutive years, starting from state j, and to be in state i at the end of the period. Eq. (4.21) and Eq. (4.22) allows us to obtain the expression that will be useful in Eq. (4.26) and Section 5.2

$$\begin{aligned} \mathrm{P}\left[\,\Phi_{t-r} = j\,,\, \Psi_t = r \,|\, \Phi_t = i\,\right] \;\; &= \;\; {}^R_r\mathrm{P}_{ji}(t)\,{}^S_r\mathrm{P}_i(t) \\ &= \;\; \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right] / \mathrm{P}\left[\,\Phi_t = i\,\right] \\ &= \;\; \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right] / p_t^{(i)}, \end{aligned} \tag{4.23}$$

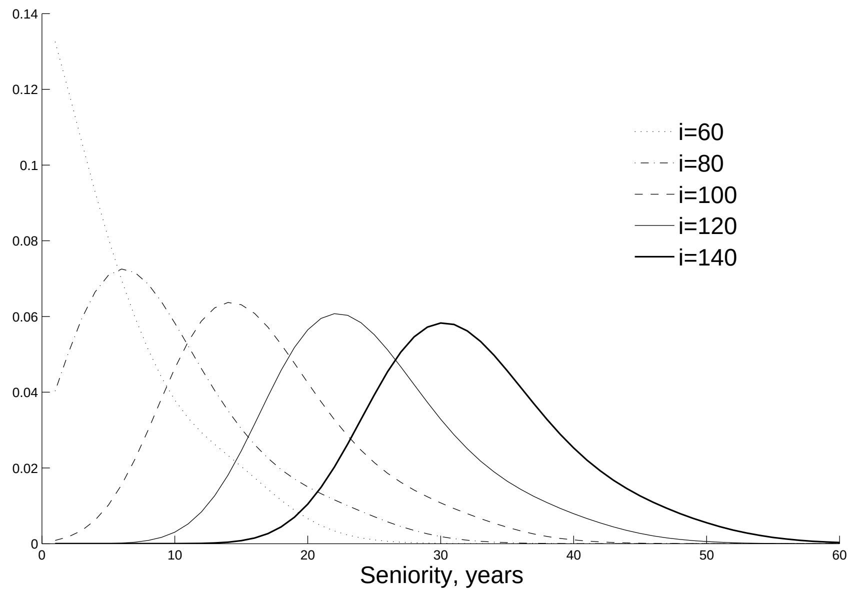

Figure 4.8: Conditional seniority distribution, ${}^{S}_{r}\mathrm{P}_i(t)$, $t = 100$, SW1911M

so we obtain

$$\mathrm{P}\left[\mathcal{E}^{(r)}_{ji}(t)\right] = {}^{R}_{r}\mathrm{P}_{ji}(t){}^{S}_{r}\mathrm{P}_i(t)p^{(i)}_t. \tag{4.24}$$

We provide an illustration of Eq. (4.22) in Fig. 4.9. On part (B) we represent schematically the event $\mathcal{E}^{(r)}_{ji}(t)$ and on part (A) we indicate that the state at time $t - r + 1$ is chosen with distribution $\mathbf{M}_{t-r+1}$.

We illustrate the reversal probabilities ${}^{R}_{r}\mathrm{P}_{ji}(t)$, $r = 0$-100, for a participant in state $i = 100$ at time $t = 100$ on the top figure in Fig. 4.10. One reads the figure from the right to the left. The right graph is the conditional distribution of the health states, j, at time 100, given zero seniority and $i = 100$ at this time. Obviously, all probability mass is concentrated at state $j = 100$. Other distributions correspond to different values of r, which increases up to 100 in the left direction. For example, the first distribution to the left of the right peak is the health state distribution of the participant, given that he/she entered the pension plan just a couple of years ago. The distribution is concentrated on values not much smaller than 100 because in a few years the state may not have changed much. The reversal probability implies that the participant is active at time t and was active r years ago, therefore, for high values of r the health state distribution at the entrance to the plan is concentrated around the youngest possible health states. Clearly, the youngest possible states

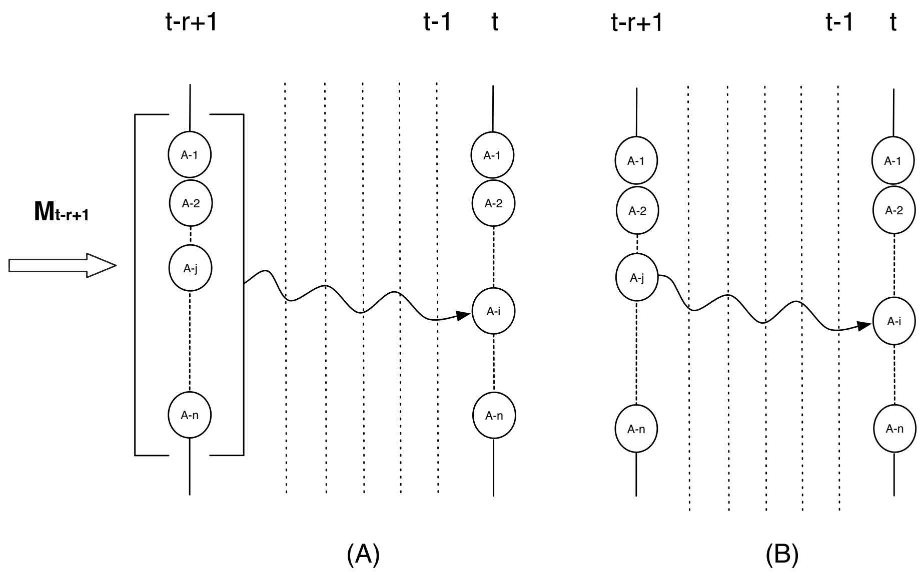

Figure 4.9: **(A)**: ${}^{r}N_t^{(i)}$, **(B)**: ${}^{R}_{r}\mathrm{P}_{ji}(t)\;{}^{r}N_t^{(i)}$, $i, j \in A$

are determined by initial health state distribution $\underline{p}_0$, that we demonstrate on the bottom figure (it is shown also in Fig. 4.2). By examining the two figures carefully, one notices that the two peaks on the left of the top figure correspond to the two left peaks in the bottom figure, amplified by the fact that we consider the distribution for $r = 100$. Obviously, given that a participant survives for $r = 100$ years, one may expect that his/her state at the entrance to the pension plan is the youngest one.

Cash inflows and outflows. Assume that all the contributions are paid at the beginning of each year. Let us denote by $F_t^+(i)$ the expected cash inflow coming at time t from participants in health state i, $i \in A$. For a participant in state i at time t we define $F_t^+(i)$ as the expected contribution: if the participant has zero service in the plan, then the contribution is given by $\Theta_t^{(i)}\mu(i)$; if the participant has r years of service, then the contribution is given by $\Theta_t^{(j)}\mu(j)$, where j is the health state at the entrance to the plan. Thus,

$$F_t^+(i) = N^o \mathrm{P}[\,\Phi_t = i\,]\left[{}^{S}_{0}\mathrm{P}_i(t)\;\Theta_t^{(i)}\mu(i) + \sum_{r=1}^{t} {}^{S}_{r}\mathrm{P}_i(t) \sum_{j\in A} {}^{R}_{r}\mathrm{P}_{ji}(t)\;\Theta_t^{(j)}\mu(j)\right],$$

N^o is the total number of participants. We use Eq. (4.19) and (4.21) to show that

$$F_t^+(i) = N^o \sum_{r=0}^{t} \sum_{j\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right]\Theta_t^{(j)}\mu(j). \tag{4.25}$$

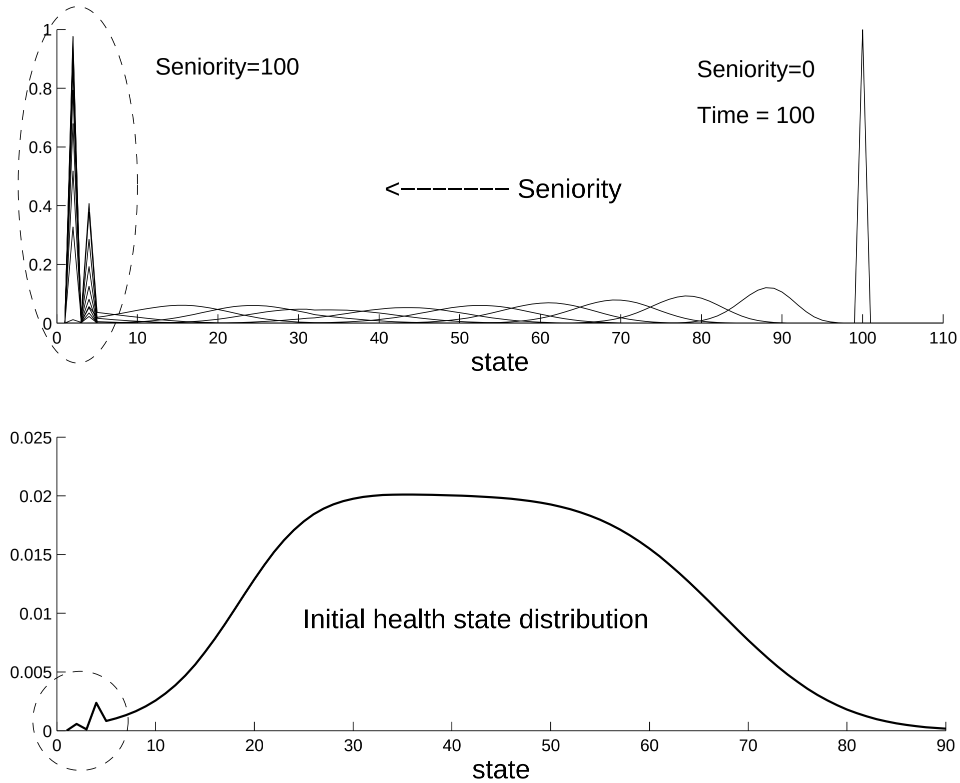

Figure 4.10: Reversal probabilities ${}^{R}_{r}\mathrm{P}_{ji}(t)$ and initial health state distribution $\underline{p}_0$
Parameters: $t = 100$, $i = 100$, $r =$ 0-100, SW1911M, $x_l = 10$, $x_u = 30$

The total cash inflow in year t is $F_t^+ = \sum_{i \in A} F_t^+(i)$.

The expected cash outflow $F_t^-(i)$ coming at time t from participants in state i, $i \in A$, consists of several terms:

- payments of lifetime annuities $\ddot{a}^{(i)}$ (see Eq. (2.12)) to participants who retire in year $t + 1$;
- periodic expenditures for current active policies, c;
- periodic plus initial expenditures for new policies, $c + I_0$;
- payments in case of death or early removal, which are equal to the expected number of contributions $\Theta_t^{(i)}\mu(i)$ paid to the plan

Thus,

$$\begin{aligned}F_t^-(i) \;=\;& \ddot{a}^{(i)} N^o \mathrm{P}[\,\Phi_t = i\,]\left(e^{\Pi}\right)_{iR} + cN^o\mathrm{P}[\,\Phi_t = i\,] + N^o(c+I_0)M_t^{(i)} + \\ & N^o\mathrm{P}[\,\Phi_t = i\,]\left(\left(e^{\Pi}\right)_{iS} + \left(e^{\Pi}\right)_{iD}\right)\sum_{r=0}^{t} {}_r^S\mathrm{P}_i(t)\sum_{j\in A} r\,{}_r^R\mathrm{P}_{ji}(t)\,\Theta_t^{(j)}\mu(j) \\ =\;& \ddot{a}^{(i)} N^o \mathrm{P}[\,\Phi_t = i\,]\left(e^{\Pi}\right)_{iR} + cN^o\mathrm{P}[\,\Phi_t = i\,] + N^o(c+I_0)M_t^{(i)} + \\ & N^o\left(\left(e^{\Pi}\right)_{iS} + \left(e^{\Pi}\right)_{iD}\right)\sum_{r=0}^{t} r\sum_{j\in A}\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right]\Theta_t^{(j)}\mu(j),\end{aligned} \tag{4.26}$$

Finally, we use Eq. (3.34) to compute the present value of the cash flow at time t:

$$PV_t = \left(F_t^+ - F_t^-\right)v^k,$$

where v is a *discount factor*.

In order to verify if the numerically obtained results make sense we perform a simple verification for F_t^- and F_t^+.

<u>Verification of F_t^+ :</u>

If we assume premiums and salaries to be equal to one for all plan participants and the interest rate to be $\delta = 0$, then the cash inflow at time t is exactly equal to the number of active plan participants in the plan.

<u>Verification of F_t^- :</u>

Under the same assumptions as for the verification of F_t^+, we can exclude the seniority and reversal probabilities from Eq. (4.26) by calculating $\hat{r}_t$, the *aggregate average seniority* in the pension plan at time t, conditional on the past being active. It is given by

$$\hat{r}_t = \frac{\sum_{r=0}^{t}\sum_{i\in A} {}^r N_t^{(i)} r}{\sum_{r=0}^{t}\sum_{i\in A} {}^r N_t^{(i)}} = \frac{\sum_{r=0}^{t} r\sum_{i\in A}\sum_{j\in A}\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right]}{\sum_{r=0}^{t}\sum_{i\in A}\sum_{j\in A}\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t)\right]}. \tag{4.27}$$

Aggregate average seniority $\hat{r}_t$ is a useful quantity. It is a general characteristic of the pension fund, which can be easily computed from individual data on plan participants at the moment of leaving the plan. Therefore, the comparison of the value obtained from Eq. (4.27) to the observed value of $\hat{r}_t$ can serve as a goodness-of-fit criterion of the model. Also, due to its independence on the current state of a participant, $\hat{r}_t$ allows us to approximate the cash outflow given by Eq. (4.26) as follows

$$\begin{aligned}\tilde{F}_t^-(i) \simeq\;& \ddot{a}^{(i)} N^o \mathrm{P}[\,\Phi_t = i\,]\left(e^{\Pi}\right)_{iR} + cN^o\mathrm{P}[\,\Phi_t = i\,] + N^o(c+I_0)M_t^{(i)} + \\ & + N^o\mathrm{P}[\,\Phi_t = i\,]((e^{\Pi})_{iS} + (e^{\Pi})_{iD})\hat{r}_t.\end{aligned}$$

Here, we also assumed that premiums and salaries to be equal to one for all plan participants. The verification tests for F_t^+ and F_t^+ are successfully performed.

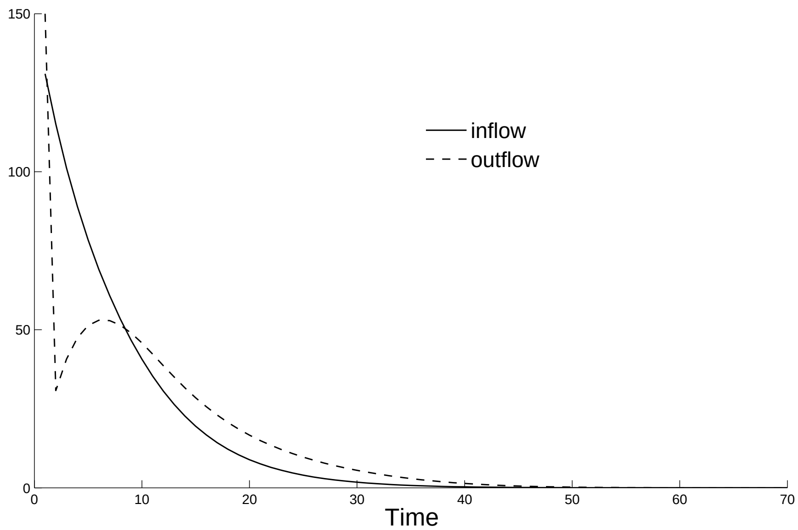

Figure 4.11: Cash flows, F_t^- and F_t^+, no replacements
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.15$, $N^o = 1000$, SW1911M

4.4 Optimal modifications

The only one parameter of the generator Π which may be chosen with a lot of freedom is ϕ, the speed at which participants are replaced in the plan after they die, retire or surrender (see Eq. (4.1)). The choice of this parameter affects the cash flows in an expected manner, as we show in Fig. 4.11 and Fig. 4.12. Due to the generator property of Π, all the characteristics of the plan will stabilize, so do the cash flows, F_t^- and F_t^+. If we choose ϕ to be close to zero, then there will be hardly any active participants in the future and, therefore, both F_t^- and F_t^+ will converge to zero, as demonstrated in Fig. 4.11. If ϕ is a positive constant, then F_t^- and F_t^+ converge to some non-zero quantity. In the illustration given in Fig. 4.12 we chose ϕ to balance future cash flows. In both figures one may notice a high value of the cash outflow in the first year and then a big drop of it in the next year. This is caused by the lump-sum initial expenditures per policy-in our model, we assume that at time zero there are N^o new participants at once, which leads to the expenditures of $I_0 N^o$ in the first year.

In the figures, we use the lump-sum pension amount 1 instead of $\ddot{a}_i^{[\theta]}$, the salaries $\Theta_t^{(i)}$, $i \in A$ are considered to be 1 as well.

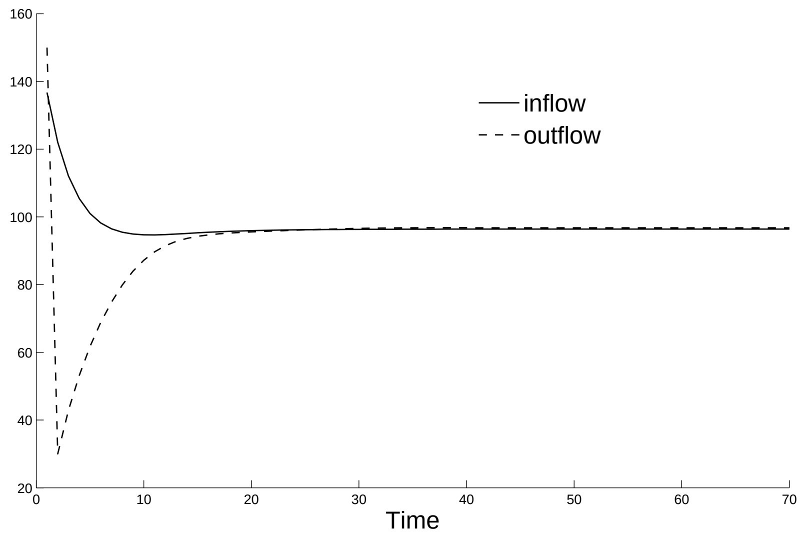

Figure 4.12: Cash flows, F_t^- and F_t^+, optimal replacements
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.15$, $N^o = 1000$, SW1911M

We examine the impact of ϕ on the future cash flows by solving the equation

$$\widehat{F}^+(\phi) = \widehat{F}^-(\phi), \tag{4.28}$$

where $\widehat{F}^+, \widehat{F}^-$ are the stationary values of the cash inflows and outflows and $1/\phi$ is the average time of the replacement for surrendered, retired and deceased plan participants. In order to calculate $\widehat{F}^+$, $\widehat{F}^-$ we obtain the stationary characteristics of the population of plan participants by assuming the initial health state distribution to be the stationary distribution of the population, given in Eq. (1.19).

We solve Eq. (4.28) numerically. The resulting difference in cash flows is presented in Fig. 4.13 for different values of $1/\phi$, the expected delay of the replacement. The Y-axis is the difference between cash inflow and cash outflow. The starred part of the curve corresponds to a positive total cash flow and the solid line corresponds to its negative values. In this example all salaries are fixed and equal to one. The behavior of the curve is quite logical – the faster the replacement, the greater the money received by the pension fund. One can see from the figure that the replacement should happen within about 3.3 years in order to have a positive long-term profit.

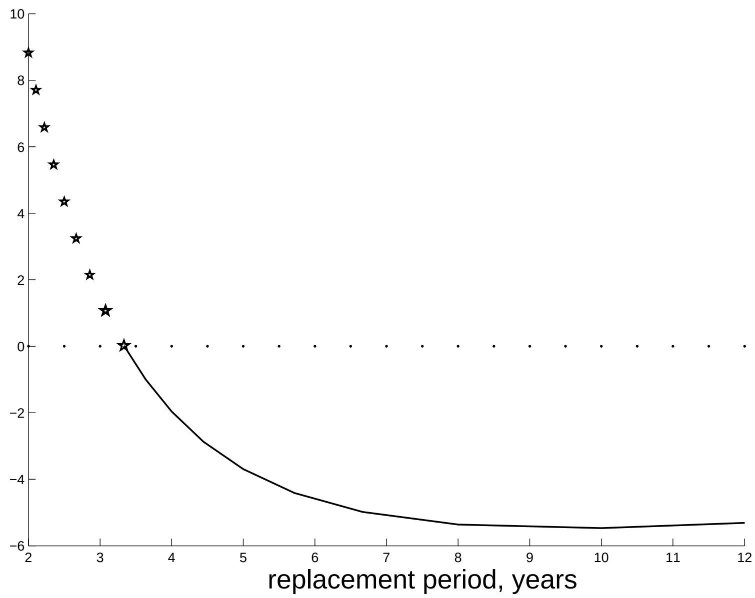

Figure 4.13: Expected stationary Cash Flow as a function of $1/\phi$
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.15$, $N^o = 1000$, SW1911M

Chapter 5

Risk management

The chapter is devoted to the risk analysis of the pension system constructed in Chapter 4. To simplify the discussion and the interpretation of the results, we set all salaries $\Theta_t^{(i)}$ to be equal to one. In this case, the contribution of a plan participant, who is in state i at the start of his pension plan, is $\mu(i)$ (see Eq. (4.16)). Furthermore, we assume that all plan participants behave independently from each other.

In Section 5.1 we state our general remarks regarding the procedure of risk assessment. In Section 5.2 we estimate a risk related to the present value of the cash flows (abbreviated as "profits and losses", or "P&L" in this chapter) with no investment benefits; this risk is caused by the independence of the survival of plan participants and we address it as *underlying risk*. For the purpose of longevity risk estimations, in Section 5.3 we perform the perturbation analysis of P&L with respect to future mortality rates.

Section 5.4 is devoted to the market risk. Here we consider two new other definitions of P&L: one with IB on the accumulated contributions, and the other one with IB on the accumulated contributions and the previous IB. We also give a numerical example to demonstrate the impact of the market assumptions on the future P&L. We devote Section 5.5 to the analysis of the pension system in the long term.

5.1 General remarks

Our aim is to evaluate different risks associated with the pension plan introduced in Chapter 4. First of all, we need to specify a risk measure and a profit and loss process.

One of the most popular applied risk measures is *value-at-risk* (abbreviated below as "VaR"). Duffie and Pan [21] define the VaR as follows. For a given time horizon t and confidence level β, the value-at-risk can be defined as the loss in market value over the time horizon t that is exceeded with probability $1-\beta$. The values of t and β are often chosen as 1 year and 99.5%, respectively. If we denote by X_t the VaR at time t and by W_t the random variable, representing the profit and loss process at time t, then X_t is such that

$$\mathrm{P}[\, W_t > X_t \,] \geq \beta. \tag{5.1}$$

The VaR is broadly used in risk-controlling by many financial institutions. Also, the regulatory environment requires financial organizations to develop proprietary risk measurement models, where VaR plays a key role.

According to Artzner et al. [3], the VaR measure exhibits some imperfections. For example, it does not satisfy the subadditivity property. This implies that if a company is forced to meet a requirement of extra capital, which does not satisfy this property, the company might be motivated to split up into two separate affiliates, which is of concern for regulators.

Despite the imperfections, the VaR measure allows us to demonstrate the ideas of risk valuation using a phase-type approach. With our purpose being to examine the whole financial result of the pension fund in year t, we define W_t to be the total profit and loss in this year and $\beta = 95\%$. Thus, in our examples X_t is treated as the threshold value of P&L in year t, exceeded with probability 95%. Obviously, the higher the X_t, the better the financial result.

We vary the exact definition of W_t depending on different circumstances and on the risk to be evaluated. In the following subsections we consider the following cases:

I: **Pure P&L**. We use this definition to examine pure profits and losses of the pension plan, for constant and changed mortality rates. The P&L are considered

- **(a)** In year t;
- **(b)** Accumulated by year t;

II: **Pure P&L in year t plus IB on contributions**. With this definition we aim to estimate the effect of the market on future P&L, which we assume to include investment benefits in order to better capture market dependence. Specifically, we suppose that the accumulated contributions by a plan participant bring an additional investment income, which depends on the state of the market. Thus, the total P&L value includes total investment income and pure P&L, defined by **I(a)**. It is natural to assume that pure P&L in year t and the IB obtained on the accumulated contributions by year t are dependent: this is due to the fact that both quantities are related to the same plan participants. However, from the computational point of view, it might be useful to examine the situation when they are independent. Thus, we obtain the total P&L

(a) Assuming that profits and investments are dependent;

(b) Assuming that profits and investments are independent;

III: Pure P&L during year t plus IB on contributions plus IB on accumulated IB. We reinforce the market dependence introduced in **II(a)** by assuming that there is an additional interest obtained on the previous investment income.

5.2 Underlying risk

We determine the distribution of P&L and the associated VaR of the pension plan with a given transition rate matrix Π (see Eq. (4.1)). According to the classification provided in the beginning of this section, we examine the P&L realized in year t first, and the P&L accumulated by year t next.

I(a): Pure P&L during year t. Denote by $W_t^{\mathbf{I(a)}}$ the P&L limited to the contributions and expenditures of the pension plan during year t and denote it by $W_t^{\mathbf{I(a)}}$. We decompose it as the sum of independent and identically distributed random variables $w_t^{\mathbf{I(a)}}$, which correspond to the plan participants, who are supposed to behave independently. Therefore, if the number of participants is big enough, we can apply the central limit theorem and conclude that the distribution of $W_t^{\mathbf{I(a)}}$ is approximately normal with mean $N^o E[w_t^{\mathbf{I(a)}}]$ and variance $N^o Var[w_t^{\mathbf{I(a)}}]$. Recall that N^o is the total number of participants.

Therefore, the problem is reduced to finding the distribution and the first moments of $w_t^{\mathbf{I(a)}}$. We find this distribution from Table 5.1, where we present four possible types of events to characterize the cost distribution of a participant at time t, along with the corresponding P&L value and probability. The table should be considered for all $i, j \in A$. The first event implies either being active in the beginning of the year and staying active for one year, or being a new participant in year t. In this case, the participant brings one discounted contribution during year t, and the

Event	P&L value	Probability
Active at the end of year t; entry state is j.	$\mu(j)v^t$	$\sum_{i\in A}\sum_r \mathrm{P}\left(\mathcal{E}_{ji}^{(r)}(t-1)\right)\left(1-(e^{\Pi})_{iQ}\mathbf{1}\right)$ $+\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{ij}$
Retires during year t; active in state i at the beginning of year t.	$-\ddot{a}^{(i)}v^t$	$p_{t-1}^{(i)}(e^{\Pi})_{iR}$
Active at the beginning of year t; dies or surrenders during year t; seniority is r, entry state is j.	$-r\mu(j)v^t$	$\sum_{i\in A}\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]\left((e^{\Pi})_{iD}+(e^{\Pi})_{iS}\right)$
Not in the plan neither at the beginning, nor at the end of year t	0	$\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{iQ}\mathbf{1}$

Table 5.1: Individual P&L , $w_t^{\mathbf{I(a)}}$

contribution depends on the state j of the participant at the time of entrance. In the second case, if a plan participant retires during year t in state i, the loss value equals the discounted lifetime annuity calculated for health state i (see Eq. (2.12)). The probability of this event is the probability $p_{t-1}^{(i)}$ to be in state i at time $t-1$ times the probability $(e^{\Pi})_{iR}$ to retire in one year from state i.

In the third case, if an active participant with seniority r and entrance health state j dies or surrenders during year t, the loss value is equal to discounted, accumulated for r years contributions. The loss value does not depend on health state at time $t-1$, therefore the probability of this event is the sum for all active states i of the probabilities of $\mathcal{E}_{ji}^{(r)}(t-1)$ times the probability to die or surrender in one year from state i. The value of $\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]$ is computed in Eq. (4.24).

Obviously, non-existing participants bring zero profit, and that is the fourth case. One may verify that the sum of all the probabilities in Table 5.1 is equal to 1.

We obtain the distribution function of $w_t^{\mathbf{I(a)}}$ numerically, first by sorting the P&L values in increasing order, then by summing up the probabilities for the same P&L values. Upon the construction of the distribution function we compute the VaR, which we illustrate in Fig. 5.1 for different t. We explain the figure by examining the evolution of the expected number of plan participants, which is given in Fig. 5.2. Note that in order to better visualize the VaR dynamics, in Fig. 5.1 as well as in our

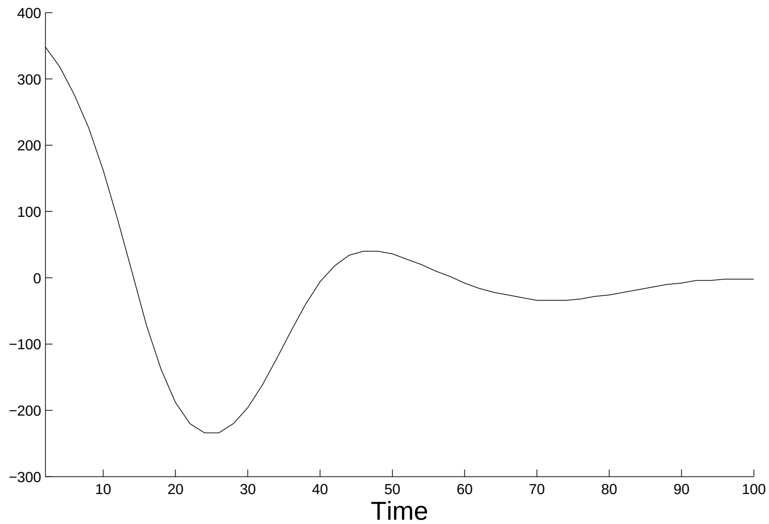

Figure 5.1: 95%-VaR, **I(a)**
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, SW1911M

further figures we cancelled the surrender option.

One observes from the figures that the VaR and the number of active plan participants evolve in a similar fashion. The initial population for the models was taken uniformly distributed between the ages 30 and 50, which implies that within 35 years all of them will be retired, as the pension age is assumed to be 65. However, those who retire or die become replaced by new participants only after some delay. Therefore, during "period of retirement", the probability of pension payment, that corresponds to the third event in Table 5.1, grows, and the probability to obtain new contributions is relatively small. This leads to the decrease of the VaR, which one can see in the interval 20-30 years. After this period, there are mostly new participants in the plan and the probability of pension payments is becoming smaller, whereas the probability of new contributions is becoming relatively high. The corresponding increase of the VaR can be seen in the period from 25 to 50 years. After this, the members of the second cohort start to retire, etc. The amplitude of these waves become smaller in time due to the convergence of the pension system to a stationary regime, reinforced by the decrease of the discount factor.

I(b): Pure P&L accumulated by year t. Denote by $W_t^{\mathbf{I(b)}}$ the P&L accumulated in the pension fund during the interval $[0, t]$. To obtain its distribution we use a similar method as for $W_t^{\mathbf{I(a)}}$. An individual P&L at time t $w_t^{\mathbf{I(b)}}$ is equal to

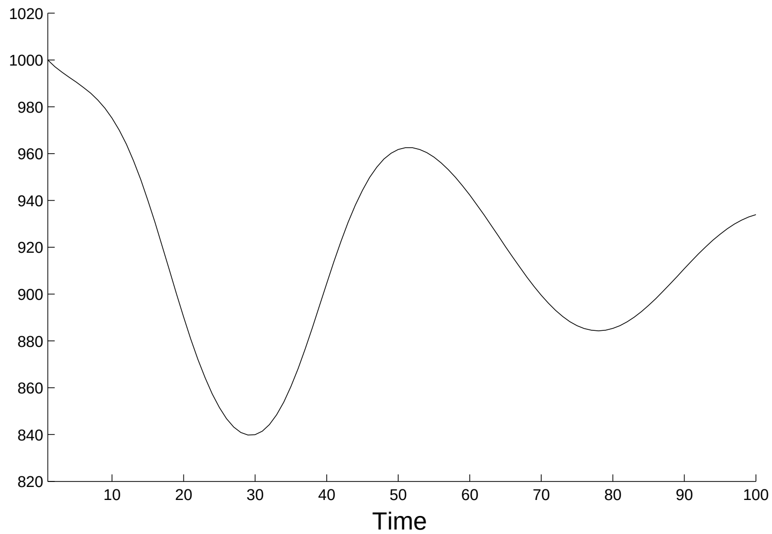

Figure 5.2: Number of active plan participants,
$N^o = 1000$, SW1911M.

the accumulated contributions, if the individual at time t is active and has a positive seniority; it is equal to the accumulated contributions minus the lifetime annuity, if the individual retires during year t; it is equal to zero, if the individual is not in the plan, or if he/she has just arrived to the plan and has zero seniority, or if the individual dies or surrenders during year t (as in this case the benefit is equal to the accumulated contributions). The distribution of $w_t^{\mathbf{I(b)}}$ is given in Table 5.2. In the same manner as for $\mathbf{I(a)}$, we approximate the distribution of $W_t^{\mathbf{I(b)}}$ by a normal distribution with mean $N^o E[w_t^{\mathbf{I(b)}}]$ and variance $N^o Var[w_t^{\mathbf{I(b)}}]$.

We illustrate the evolution of VaR in Fig 5.3. Here as before, we chose the initial population to be uniformly distributed between the ages 30 and 50, and the pension age is assumed to be 65. The VaR increases almost linearly during the first 15 years, which is reasonable because the probability of retirements and, therefore, of pension payments is rather small. Between the 15th and 35th year from the start of the plan all initial participants become retired. Due to this reason and due to a delay of the replacement we see a local minimum of VaR in year 35. After this year there are mostly new participants in the pension system, therefore, accumulations increase, so does VaR. And again, in 30 years new plan participants begin to retire, etc. In the long-term perspective the probabilities will stabilize and VaR will be decreasing due to the discount factor.

Event	P&L value	Probability
Active at the end of year t; entry state is j; seniority $r > 0$	$r\mu(j)v^t$	$\sum_{i\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]\left(1-(e^{\Pi})_{iQ}\mathbf{1}\right)$
Active in the end of year t; entry state is j; seniority $r = 0$	0	$\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{ij}$
Active at the beginning of year t; retires during year t; current state i; seniority r; entry phase is j	$r\mu(j)v^t - \ddot{a}^{(i)}v^t$	$\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right](e^{\Pi})_{iR}$
Active in the beginning of year t; seniority r; entry state is j; death or surrender in year t	0	$\sum_{i\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]\left((e^{\Pi})_{iD}+(e^{\Pi})_{iS}\right)$
Not in the plan neither at the beginning, nor at the end of year t	0	$\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{iQ}\mathbf{1}$

Table 5.2: Individual P&L , $w_t^{\mathbf{I(b)}}$

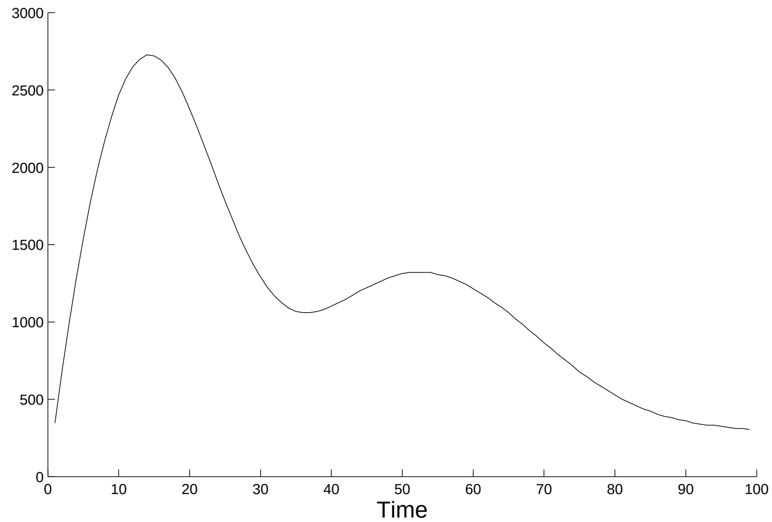

Figure 5.3: 95%-VaR, **I(b)**
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, SW1911M

5.3 Longevity risk

In this subsection we apply one of the methods that we suggested in Section 2.4 to estimate the *longevity risk*. Specifically, in this example we assume that the decrease of the mortality rates, starting from some time τ_m, is caused by environmental factors. Thus, we analyze the P&L in a perturbed environment. The environment in the pension fund is given by the generator Π (see Eq. (4.1)), which contains mortality rates for each active state. One may consider different structures of the perturbation of the rates, depending on the pursued objectives; here, we assume that, starting from some time τ_m, the mortality rates become smaller for all health states.

We apply the technique, described in Section (2.4), particularly, in Eq. (2.29) and define the generator of the perturbed pension system as

$$\tilde{\Pi} = \Pi + \varepsilon \hat{D}(\underline{q}), \tag{5.2}$$

where ε is a positive scalar and

$$\hat{D}(\underline{q}) = \begin{pmatrix} D(\underline{q}) & \mathbf{0} & -\underline{q} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{5.3}$$

Here, $D(\underline{q})$ is the diagonal matrix with vector $\underline{q}$ on the diagonal. The distribution of health states at time t, $\underline{\tilde{p}}_t = (\tilde{p}_t^{(i)} : i = 1, \ldots, n+3)$, is given by

$$\underline{\tilde{p}}_t = \begin{cases} \underline{p}_0 e^{\Pi t}, & t < \tau_m \\ \underline{p}_0 e^{\Pi \tau_m} e^{\tilde{\Pi}(t-\tau_m)}, & t \geq \tau_m \end{cases} .$$

The seniority distribution vector for the perturbed system is ${}^r\tilde{\mathbf{N}}_t = \left({}^r\tilde{N}_t^{(i)},\ i \in A\right)$, where ${}^r\tilde{N}_t^{(i)} = \mathbf{P}\,[\,\Psi_t = r\,,\, \Phi_t = i\,]$ is the probability that a participant at time t has physiological age i and seniority r in the plan. The seniority distribution depends on the time when the perturbation happens. From Eq. (4.20) and Eq. (4.4) we find that for all $r < t$:

- if $t < \tau_m$, then
$$ {}^r\tilde{\mathbf{N}}_t = \underline{\tilde{p}}_{t-r}^{(Q)} \left(e^{\Pi}\right)_{(Q,A)} e^{\Pi_{AA}(r-1)};$$
- if $(t - r + 1) \leq \tau_m \leq t$, then
$$ {}^r\tilde{\mathbf{N}}_t = \underline{\tilde{p}}_{t-r}^{(Q)} \left(e^{\Pi}\right)_{(Q,A)} e^{\Pi_{AA}(\tau_m-(t-r+1))} e^{\tilde{\Pi}_{AA}(t-\tau_m)};$$
- if $\tau_m \leq t - r$, then
$$ {}^r\tilde{\mathbf{N}}_t = \underline{\tilde{p}}_{t-r}^{(Q)} \left(e^{\tilde{\Pi}}\right)_{(Q,A)} e^{\tilde{\Pi}_{AA}(r-1)}.$$

Here, $\tilde{\Pi}_{AA} = \Pi_{AA} + \varepsilon D(\underline{q})$ is the matrix that describes only active participants of the perturbed model.

$$ {}_r^S\tilde{\mathbf{P}}_i(t) = {}^r\tilde{N}_t^{(i)} / \tilde{p}_t^{(i)},$$

for the perturbed conditional probabilities ${}_r^S\tilde{\mathbf{P}}_i(t) = \mathrm{P}\,[\,\Psi_t = r\,|\,\Phi_t = i\,]$ for active plan participants.

The various conditional reversal probabilities ${}_r^R\tilde{\mathbf{P}}_{ji}(t)$ are obtained exactly like at the beginning of Section 4.3 and we find that

- if $t < \tau_m$, then
$$ {}_r^R\tilde{\mathbf{P}}_{ji}(t) = \underline{\tilde{p}}_{t-r}^{(Q)} \left(e^{\Pi}\right)_{(Q,j)} \left(e^{\Pi_{AA}(r-1)}\right)_{ji} / \ {}^r\tilde{N}_t^{(i)} \ ;$$
- if $(t - r + 1) \leq \tau_m \leq t$, then
$$ {}_r^R\tilde{\mathbf{P}}_{ji}(t) = \underline{\tilde{p}}_{t-r}^{(Q)} \left(e^{\Pi}\right)_{(Q,j)} \left(e^{\Pi_{AA}(\tau_m-(t-r+1))} e^{\tilde{\Pi}_{AA}(t-\tau_m)}\right)_{ji} / \ {}^r\tilde{N}_t^{(i)} \ ;$$

Event	P&L value	Probability
Active at the end of year t; entry state is j.	$\mu(j)v^t$	$\sum_{i\in A}\sum_r \mathrm{P}\left[\tilde{\mathcal{E}}_{ji}^{(r)}(t-1)\right]\left(1-(e^{\tilde{\Pi}})_{iQ}\mathbf{1}\right)$ $+\sum_{i\in Q}\tilde{p}_{t-1}^{(i)}(e^{\tilde{\Pi}})_{ij}$
Retires during year t; active in state i at the beginning of year t.	$-\tilde{\ddot{a}}^{(i)}v^t$	$\tilde{p}_{t-1}^{(i)}(e^{\tilde{\Pi}})_{iR}$
Active at the beginning of year t; dies or surrenders during year t; seniority is r, entry state is j.	$-r\mu(j)v^t$	$\sum_{i\in A}\mathrm{P}\left[\tilde{\mathcal{E}}_{ji}^{(r)}(t-1)\right]\left((e^{\tilde{\Pi}})_{iD}+(e^{\tilde{\Pi}})_{iS}\right)$
Not in the plan neither at the beginning, nor at the end of year t	0	$\sum_{i\in Q}\tilde{p}_{t-1}^{(i)}(e^{\tilde{\Pi}})_{iQ}\mathbf{1}$

Table 5.3: Individual P&L , $\tilde{w}_t^{\mathbf{I(a)}}$

- if $\tau_m \leq t-r$, then

$$ {}^R_r\tilde{\mathbf{P}}_{ji}(t) = \underline{\tilde{p}}_{t-r}^{(Q)}\left(e^{\tilde{\Pi}}\right)_{(Q,j)}\left(e^{\tilde{\Pi}_{AA}(r-1)}\right)_{ji} / \, {}^r\tilde{N}_t^{(i)} \, . $$

The probability of event $\mathcal{E}_{ji}^{(r)}(t)$, which means that at time t a plan participant is in state i, $i \in A$ and he/she joined the pension plan at time $t-r$ being in state j, $j \in A$ (see Eq. (4.18)), is given by the expression (4.24), where one has to replace the seniority and reversal probabilities by their perturbed values.

Define by $\widetilde{W}_t^{\mathbf{I(a)}}$ and $\tilde{w}_t^{\mathbf{I(a)}}$ the perturbed versions of $W_t^{\mathbf{I(a)}}$ and $w_t^{\mathbf{I(a)}}$, respectively. The distribution of $\tilde{w}_t^{\mathbf{I(a)}}$ does not change for $t \leq \tau_m$. For $t > \tau_m$, it is given in Table 5.3. The table is mostly identical to Table 5.1; the differences are in the column of probabilities, which are replaced by their perturbed values, computed above, and in the P&L value for the event "retirement", where $\tilde{\ddot{a}}^{(i)}$ is used instead of $\ddot{a}^{(i)}$. The quantity $\tilde{\ddot{a}}^{(i)}$ is the annuity computed with the new transition matrix.

Like in case $\mathbf{I(a)}$, the large number of independent plan participants allows one to conclude that $\tilde{W}_t^{\mathbf{I(a)}}$ has approximately normal distribution with mean $N^o E[\tilde{w}_t^{\mathbf{I(a)}}]$ and variance $N^o Var[\tilde{w}_t^{\mathbf{I(a)}}]$. We illustrate one example on Fig. 5.4, where we decrease the mortality rates at time $\tau_m = 5$ years and $\varepsilon = 0.3$. The solid line represents the VaR of the original model, the dashed line represents the VaR of the perturbed

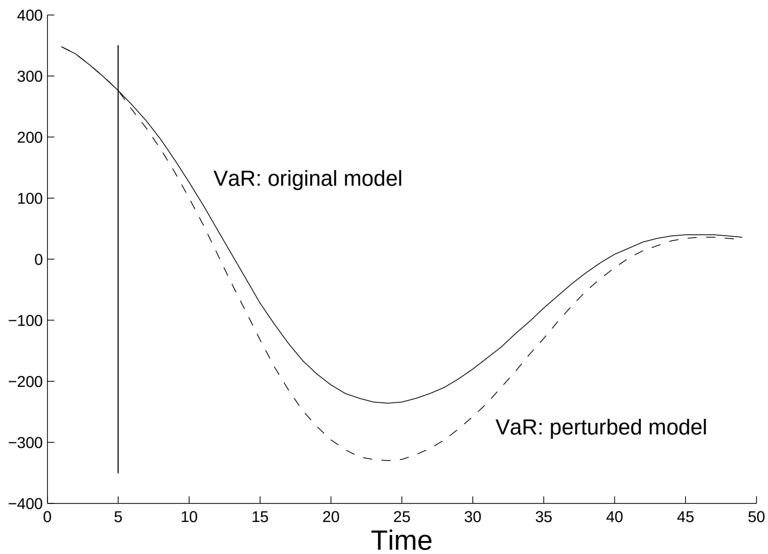

Figure 5.4: 95%-VaR, $\mathbf{I(a)}$, longevity effect
Parameters: $v = 0.971$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, $\tau_m = 5$, $\varepsilon = 0.3$, SW1911M

model. As one can observe, the VaRs are the same for $t \leq 5$; for $t > 5$, the VaR of the perturbed system is lower. The difference between the VaRs is bigger for the periods when participants mostly retire, and it is not so big when new participants appear.

We are also interested in additional questions about the perturbed pension system, that are not related to VaR; these are presented in Section 5.5.

5.4 Market risk

Here, we need a model for the market and we adopt the approach introduced in Norberg [46], where the economical environment is described by a *Markov reward process* (abbreviated below as "MRP"). This MRP includes a discrete time Markov chain, and a reward function that associates a reward value for each state of the Markov chain. Formally, MRP is usually defined as a triplet $(S, \underline{\delta}, P)$, where S is a finite or countable state space, $\underline{\delta}$ is a vector that contains a reward value for each state in S, and P is a transition probability matrix.

Here, we take $S = \{1, ..., m\}$ and $\underline{\delta} = \left(\delta^{(1)}, .., \delta^{(m)}\right)$ with the interpretation that δ_t is the investment rate process at time t. It can take one of m values and is such that

$\delta_t = \delta^{(Y_t)}$, where Y_t is the state of the Markov chain at time t. P is the transition probability matrix of Y_t.

Let $\underline{y}_0$ be the initial probability vector. Then, according to Eq. (1.20), the distribution of the market states in the end of year t is determined by

$$\underline{y}_t = \underline{y}_0 P^t. \tag{5.4}$$

II(a): Pure P&L in year t plus IB on contributions. Profits and investments are dependent.

We denote the total P&L during year t by $W_t^{\mathbf{II(a)}}$ and, as before, in order to obtain the distribution of $W_t^{\mathbf{II(a)}}$ we obtain the distribution of an individual P&L, denoted by $w_t^{\mathbf{II(a)}}$. The distribution of $w_t^{\mathbf{II(a)}}$ is given in Table 5.4. The table differs from Table 5.1 in two aspects. Firstly, we change "P&L value" column of Table 5.1 by adding the IB on accumulated contributions. The IB is added for those plan participants who are active at the beginning of year t; for plan participants with seniority r and entry state j the discounted value of IB equals $\delta r \mu(j) v^t$, where δ is the realization of the investment rate process δ_t in year t. We assume that the market and the participants are independent, so that the P&L value for each participant does not depend on the particular realization of δ_t. Secondly, due to the fact that the IB value depends on r, we need to separately consider cases $r > 0$ and $r = 0$ for participants who are active at the end of year t.

In the definition **II(a)** we say that profits and investments are "dependent", because both of them are incorporated in the individual distribution of $w_t^{\mathbf{II(a)}}$. This is not the case for the approximation that we consider in **II(b)** in the next paragraph.

Like in all previous cases, $W_t^{\mathbf{II(a)}}$ is approximately normal with the parameters $E[w_t^{\mathbf{II(a)}}(\delta)]$ and $Var[w_t^{\mathbf{II(a)}}(\delta)]$. By the law of total probability,

$$\mathrm{P}\,[\,W_t^{\mathbf{II(a)}} < x\,] = \sum_{k=1}^{m} (\underline{y}_t)_k \int_{(-\infty,x)} g(u; N^o E[w_t^{\mathbf{II(a)}}(\delta^{(k)})], N^o Var[w_t^{\mathbf{II(a)}}(\delta^{(k)})])du \tag{5.5}$$

where $g(u; m, \sigma^2)$ is the normal density function with parameters m and σ^2, evaluated at u, and $(\underline{y}_t)_k = \mathrm{P}[\,Y_t = k\,]$ is given by Eq. (5.4).

II(b): Pure P&L in year t plus IB on contributions. Profits and investments are independent. We present here an approach, which serves as an approximation of **II(a)**. Instead of tracking individual P&L in year t and computing the IB for each participant, we might lump all the contributions at the beginning of year t and compute the IB on this total amount. Then, the total P&L in year t equals pure P&L in year t plus this IB. By assuming independence between pure P&L and

Event	P&L value	Probability
Active at the end of year t; entry state is j; seniority $r > 0$.	$\mu(j)v^t$ $+\,\delta r\mu(j)v^t$	$\sum_{i\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]\left(1-(e^{\Pi})_{iQ}\mathbf{1}\right)$
Active at the end of year t; entry state is j; seniority $r = 0$.	$\mu(j)v^t$	$\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{ij}$
Retires during year t; active in state i at the beginning of year t; seniority r.	$-\ddot{a}^{(i)}v^t$ $+\,\delta r\mu(j)v^t$	$\mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right](e^{\Pi})_{iR}$
Active at the beginning of year t; dies or surrenders during year t; seniority is r, entry state is j.	$\delta r\mu(j)v^t$ $-\,r\mu(j)v^t$	$\sum_{i\in A} \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right]\left((e^{\Pi})_{iD}+(e^{\Pi})_{iS}\right)$
Not in the plan neither at the beginning, nor at the end of year t.	0	$\sum_{i\in Q} p_{t-1}^{(i)}(e^{\Pi})_{iQ}\mathbf{1}$

Table 5.4: Individual P&L , $w_t^{\mathbf{II(a)}}$, conditional on $\delta_t = \delta$.

the IB, we obtain that the total P&L is a sum of two independent random variables. We show that, despite the approximation, this method gives similar results to those of the model **II(a)**.

Denote the IB in year t by J_t and the accumulations of the fund at the beginning of year t by L_t. The IB is defined as an interest on the accumulated contributions, so that

$$J_t = L_t \delta_t v^t \qquad \forall t, \tag{5.6}$$

and the total P&L in year t is directly expressed as

$$W_t^{\mathbf{II(b)}} = W_t^{\mathbf{I(a)}} + J_t, \tag{5.7}$$

where $W_t^{\mathbf{I(a)}}$ is the P&L defined in Section 5.2. Our objective is to determine the probability density function of $W_t^{\mathbf{II(b)}}$, which be denote by $f_t^{\mathbf{II(b)}}(\cdot)$. As a first step, we obtain the distribution of L_t by applying the central limit theorem to the sum of individual accumulations, that we call l_t. This is equal to the number of accumulated contributions, if the participant is active, and it is equal to 0, otherwise. The moments of l_t are given by

$$\begin{aligned} E[l_t] &= \sum_{i \in A} \sum_{j \in A} \sum_{r \geq 0} \mu(j) r \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right], \\ Var[l_t] &= \sum_{i \in A} \sum_{j \in A} \sum_{r \geq 0} (\mu(j) r)^2 \mathrm{P}\left[\mathcal{E}_{ji}^{(r)}(t-1)\right] - (E[l_t])^2. \end{aligned} \tag{5.8}$$

Thus, the distribution of L_t is approximately normal with parameters $N^o E[l_t]$ and $N^o Var[l_t]$. We have assumed that the investment rate process is independent of the fund, so that L_t and δ_t are independent and the distribution of J_t, defined by Eq. (5.6), is given by the mixture

$$\mathrm{P}\left[\, J_t < x \,\right] = \sum_k (\underline{y}_t)_{(k)} \int_{(-\infty, x/\delta^{(k)})} g(u; N^o E[l_t], N^o Var[l_t]) du.$$

Denote the probability density function of J_t by $f_{J_t}(\cdot)$. The assumption of the independence between $W_t^{\mathbf{I(a)}}$ and J_t, and Eq. (5.7) allow to find $f_t^{\mathbf{II(b)}}(\cdot)$ as a convolution, given by

$$f_t^{\mathbf{II(b)}}(z) = \int_{(-\infty, +\infty)} g(u; N^o m_t, N^o \sigma_t^2) f_{J_t}(z-u) du. \tag{5.9}$$

The numerically computed cdfs of $W_t^{\mathbf{II(a)}}$ and $W_t^{\mathbf{II(b)}}$ are close to each other, as we show in Fig. 5.5 for t = 1-100, in dashed and solid lines. One may hardly see the

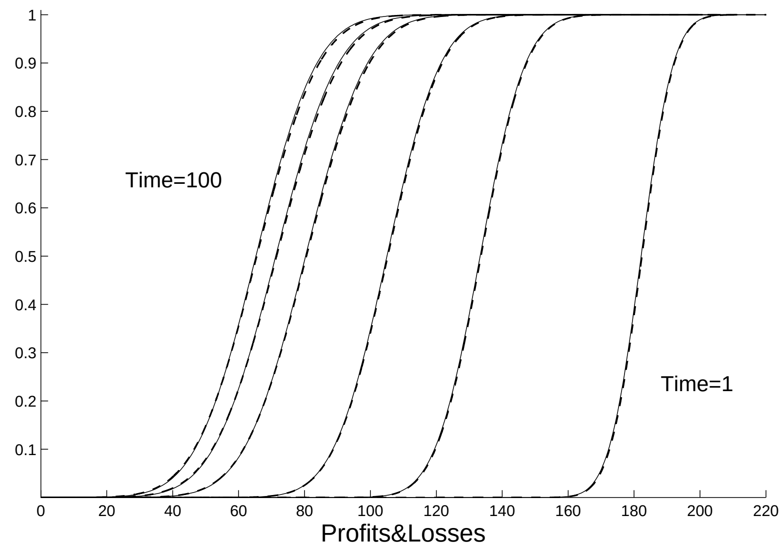

Figure 5.5: Cumulative distribution functions of $W_t^{\mathbf{II(a)}}$ and $W_t^{\mathbf{II(b)}}$
$v = 0.8$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, t = 1-100, SW1911M

difference between the lines and, therefore, conclude that the effect of the dependence between total P&L and IB is negligible.

III: Pure P&L during year t plus IB on contributions plus IB on accumulated IB. Assume that the IB obtained from the accumulated premiums in **II(a)** is itself re-invested. We are interested mostly in the effect given by the fluctuations of the investment rate; we suppose that for a long existing pension fund these fluctuations have a bigger impact than the fluctuations related to the population of pension fund participants. Thus, for the analysis we replace the total accumulations L_t of the fund by

$$L_* = \lim_{t\to\infty} E[L_t] = N^o E[l],$$

where $l = \lim_{t\to\infty} E[l_t]$. The limits exist due to the generator property of matrix Π.

We define the total IB in year t as

$$\bar{J}_t = \bar{J}_{t-1}\delta_t v + L_*\delta_t v^t, \quad \bar{J}_0 = 0, \tag{5.10}$$

where δ_t is the investment rate process, defined for model **II(a)**. Since δ_t takes one of m possible values, we can write that

$$\begin{aligned}\bar{J}_t(\delta_1, ..., \delta_t) &= L_*\delta_t v^t + L_*\delta_t\delta_{t-1}v^t + ... + L_*\delta_t \cdots \delta_1 v^t \\ &= L_* v^t \sum_{k=1}^{t} \prod_{s=k}^{t} \delta_s. \end{aligned} \tag{5.11}$$

In order to obtain the distribution of $\bar{J}_t$ we find it convenient to express it in terms of the distribution of $\bar{J}_{t-1}$. Define $H_t(x, i) = \mathrm{P}[\,\bar{J}_t \leq x,\, Y_t = i\,]$, where Y_t is the Markov chain that describes the states of the market. Define also

$$H_t(x) = \mathrm{P}[\,\bar{J}_t \leq x\,] = \sum_i H_t(x, i). \tag{5.12}$$

We obtain $H_t(x, i)$ from

$$H_t(x, i) = \int_0^\infty \sum_{j=1}^{m} dH_{t-1}(y, j)\mathrm{P}[\,\bar{J}_t \leq x,\, Y_t = i \,|\, \bar{J}_{t-1} = y,\, Y_{t-1} = j\,].$$

Here,

$$\begin{aligned}&\mathrm{P}[\,\bar{J}_t \leq x,\, Y_t = i \,|\, \bar{J}_{t-1} = y,\, Y_{t-1} = j\,] \\ &= \mathrm{P}[\,Y_t = i \,|\, Y_{t-1} = j\,] \cdot \mathrm{P}[\,\bar{J}_t \leq x \,|\, \bar{J}_{t-1} = y,\, Y_{t-1} = j,\, Y_t = i\,] \\ &= (P)_{(j,i)}\, \mathbb{1}_{\left\{(yv + L_* v^t)\delta^{(i)} \leq x\right\}}. \end{aligned} \tag{5.13}$$

Therefore,

$$\begin{aligned} H_t(x, i) &= \int_0^\infty \sum_{j=1}^{m} dH_{t-1}(y, j)\, (P)_{(j,i)}\, \mathbb{1}_{\left\{(yv + L_* v^t)\delta^{(i)} \leq x\right\}} \\ &= \sum_{j=1}^{m} H_{t-1}\left(\frac{x}{v\delta^{(i)}} - L_* v^{t-1}, j\right) (P)_{(j,i)}\,. \end{aligned} \tag{5.14}$$

The total P&L in year t is defined by

$$W_t^{\mathbf{III}} = W_t^{\mathbf{I(a)}} + \bar{J}_t,$$

where $W_t^{\mathbf{I(a)}}$ and $\bar{J}_t$ are independent for any t. If we use Eq. (5.14), (5.12), the probability density function of $W_t^{\mathbf{III}}$, denoted by $f_t^{\mathbf{III}}(z)$, is a convolution of two continuous distributions

$$f_t^{\mathbf{III}}(z) = \int_{(-\infty,+\infty)} h_t(z - u) g(u; N^o E[w_t^{\mathbf{II(a)}}], N^o Var[w_t^{\mathbf{II(a)}}]) du,$$

where $h_t(z) = dH_t(z)/dz$.

Market model examples. The incorporation of IB to P&L, given by **II(a)**, **II(b)** and **III**, increases the impact of the market on VaR. According to our assumptions, the market is modeled by a Markov reward process, namely, by the Markov Chain Y_t with a finite number of states, and by the associated with it interest rate process δ_t. We are interested in the impact of the parametrization of this process on the distribution of P&L and VaR. Below, we consider two examples.

Example 1. Constant investment rate. Assume that $\delta_t = \delta$ for all t. According to Eq. (5.6), for **II(b)**, the expected value of IB is given by

$$E[J_t] = \delta v^t E[L_t],$$

with random variable L_t being the accumulations of the fund at time t. It follows from Eq. (5.11) that the IB for case **III** is

$$\bar{J}_t \quad = v^t L_* \sum_{k=1}^{t} \delta^k = \delta v^t L_* \frac{1-\delta^t}{1-\delta}$$

Let us compare $E[J_t]$ and $\bar{J}_t$ for all t. For high t and $\delta < 1$, the ratio $E[J_t]/\bar{J}_t$ is approximately equal to $1-\delta$, where δ is the investment rate value. If δ is small enough, one can easily see that $E[J_t]$ is close to $\bar{J}_t$. For smaller values of t, prior to the *time to stability*, $E[J_t]$ is smaller than $\bar{J}_t$, because $E[L_t] < L_*$ and $(1-\delta^t)/(1-\delta)$ is greater than one.

The same relations hold for the corresponding VaRs, because the total P&L, for **II(b)** and for **III**, is equal to the sum of the P&L, defined in **I(a)**, and the IB. We illustrate the VaRs for **I(a)**, **II(b)** and **III**, for $\delta = 5\%$ in Fig. 5.6, where we depict the obtained relations. Obviously, the VaR for **I(a)** is the smallest as the corresponding P&L does not include any IB.

Another important aspect is the sensitivity of the VaRs to the discount rate v. In Fig. 5.7 we show the VaRs computed with methods **II(b)** and **III** for different values of v. As one may expect, the higher the value of v the lower the VaR. Also, the value of v does not change the relation between the two VaRs - they get closer to each other with high values of t.

Example 2. Multiple market states: same constant expected investment rate. We demonstrate the impact of the choice of the number of market states. We construct two models of the market: with different number of states, n_1 and n_2, different generator matrices, $P = e^{\Omega_1}$ and $P = e^{\Omega_2}$, but the same expected investment rate in the long term.

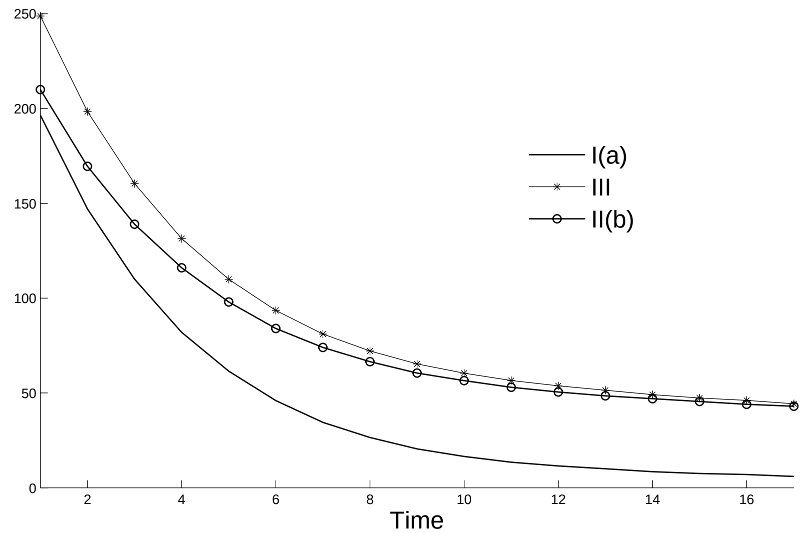

Figure 5.6: 95%-VaR, **I(a)**, **II(b)** and **III**
$m = 2$, $\delta_t = 5\% \; \forall t$, $v = 0.98$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, SW1911M

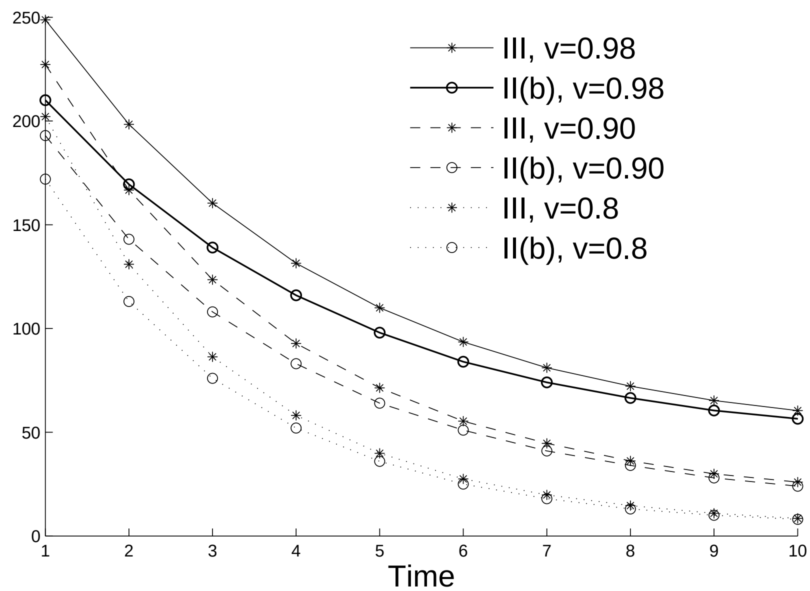

Figure 5.7: 95%-VaR **II(b)** and **III** for different v
$m = 2$, $\delta_t = 5\% \; \forall t$, $c = 0.01$, $I_0 = 0.25$, $N^o = 1000$, SW1911M

Denote by $\underline{\pi}_t^{(1)}$ and by $\underline{\pi}_t^{(2)}$ the stationary distributions of the market states for the models with n_1 and n_2 states, respectively. Denote the set of investment rate values for the first model by $\underline{\delta}_1$ and by $\underline{\delta}_2$ for the second model. We choose $\underline{\delta}_1$ and $\underline{\delta}_2$ such that

$$\underline{\pi}^{(1)}\underline{\delta}_1^{\mathrm{T}} = \underline{\pi}^{(2)}\underline{\delta}_2^{\mathrm{T}}, \tag{5.15}$$

where $\underline{\pi}^{(1)}\underline{\delta}_1^{\mathrm{T}}$ is the expected investment rate in the model with n_1 states, $\underline{\pi}^{(2)}\underline{\delta}_2^{\mathrm{T}}$ – in the model with n_2 states.

Specifically, $n_1 = 2$ and $n_2 = 3$,

$$\Omega_1 = \begin{pmatrix} -\omega & \omega \\ \omega & -\omega \end{pmatrix}, \quad \omega > 0 \tag{5.16}$$

and

$$\Omega_2 = \begin{pmatrix} -\omega & 3/4\omega & 1/4\omega \\ 3/4\omega & -\omega & 1/4\omega \\ 1/2\omega & 1/2\omega & -\omega \end{pmatrix}, \tag{5.17}$$

and

$$\delta_1^{(1)} = 1\%, \;\; \delta_1^{(2)} = 3.5\%, \tag{5.18}$$

$$\delta_2^{(1)} = 2.3\%, \;\; \delta_2^{(2)} = 3.4\%, \;\; \delta_2^{(3)} = 0\%. \tag{5.19}$$

It is easy to verify that the expected interest rate remains the same and its value is around 2.3%. In the first model, we describe a rather calm market, where the investment rate can take only two possible values with the same probability. In the second model, given by Ω_2, we allow for a rare crisis state where $\delta = 0$; the probability to leave this crisis state to one of the calm states is the same.

We illustrate the interest rate behavior for the two markets in Fig. 5.8. One can see from the figure, that the market with three states corresponds to a more risky behavior of the market.

We examine the distribution of P&L, which we present in Fig. 5.9 for $t = 10$ and $t = 25$. The two subfigures of the first column depict the P&L distribution, given by **II(b)**, for the market with two states; the two subfigures of the second column correspond to the market with three states. The top two subfigures correspond to the non-stable period: during this period P&L are affected by changes of the population of plan participants, which provokes a large variance. In the presented example, there are many surrendered participants at $t = 10$, which have not been replaced yet, therefore, the mean of the distributions is slightly negative. For the stationary period, pure profits and losses are balanced and the positive means of the distributions on the two bottom subfigures are due to the inclusion of additional IB. The variance of the bottom distributions shrinks due to the convergence of the Markov chains, for plan participants and for the market, to a stationary regime.

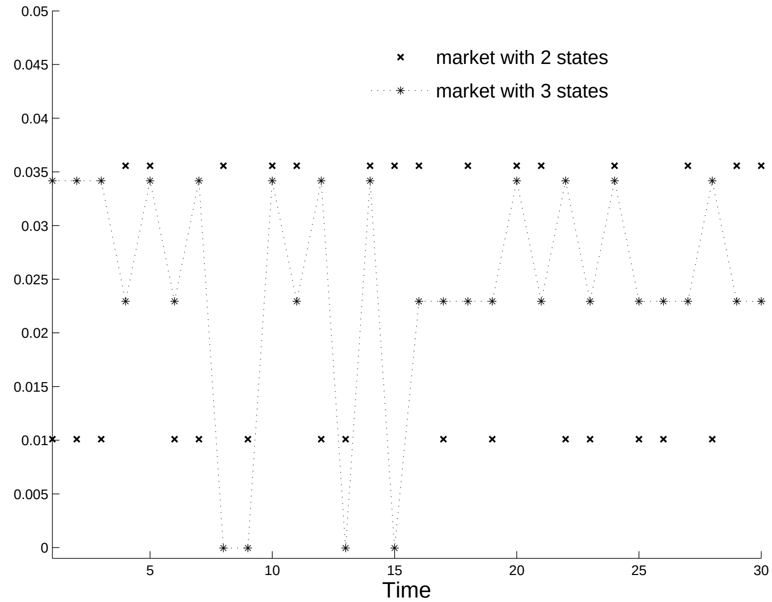

Figure 5.8: Interest rate processes
Parameters: $n_1 = 2$, $n_2 = 3$, Eq. (5.16) for Ω_1, Eq. (5.17) for Ω_2,
$\omega = 4$, Eq. (5.18) for $\underline{\delta}_1$, Eq. (5.19) for $\underline{\delta}_2$.

5.5 Stability analysis

In this section we are interested in the stability analysis of the pension system, controlled by a Markov chain with generator Π (see Eq. (4.1)).

One interesting characteristic of a Markov chain system is the time it requires before reaching stability. The *time to stability* t^* is the length of time until the characteristics of the population become stable. This is useful to know when choosing the time horizon for the cash flows calculations. Due to the properties of the transition probability matrices, the matrix $\lim_{t\to\infty} e^{\Pi t}$ has all zero eigenvalues, except one which is equal to one. The matrix e^{Π} itself has one eigenvalue which is equal to one and the others that are strictly less than one in absolute value. Taking these facts into consideration we define t^* as the smallest value such that

$$\lambda^{t^*} = \varepsilon, \tag{5.20}$$

where λ is the second maximal eigenvalue of the matrix e^{Π} and ε is the required degree of precision, so that $e^{\Pi t}$ is nearly constant for $t > t^*$.

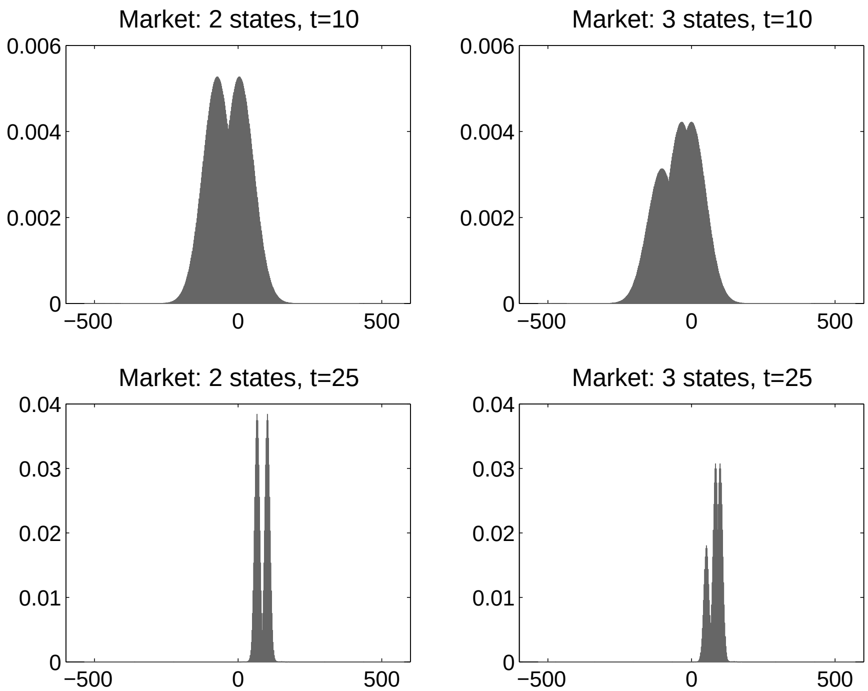

Figure 5.9: P&L distribution, **II(b)**, for two markets with same expected investment rate of 2.3%, different number of states, SW1911M

We illustrate the distribution $\underline{p}_t^{(A)}$ of health states of plan participants (4.6) calculated for t prior and after the time to stability in Fig. 5.10. The initial distribution $\underline{p}_0^{(A)}$, given by Eq. (4.9), is represented by the thin solid line, intermediate distributions are the dashed lines and the thick solid line is the stationary distribution.

Another interesting characteristic that requires a similar approach is the *maximal seniority* of a plan participant in the pension system, that is, the maximum number of years that an active plan participant may reasonable be expected to spend in the plan. Assume that the plan participant is in the state i, then the probability to remain for t years as an active participant is $\underline{\alpha}^{(i)} e^{\Pi_{AA} t}\mathbf{1}$, where $\underline{\alpha}^{(i)} = (\alpha_j^{(i)}, j \in A)$ is defined by Eq. (2.11). In order to represent the maximum service we define the time t^a such that

$$\forall t \geq t^a: \quad \max_{i \in A} \{\underline{\alpha}^{(i)} e^{\Pi_{AA} t}\mathbf{1}\} \leq \epsilon, \tag{5.21}$$

where ϵ is small enough. The problem is similar to the one for the time to stability, where the solution is given by Eq. (5.20) and $t^a = \log \epsilon / \lambda_A$ is the solution to $e^{\lambda_A t^a} = \epsilon$,

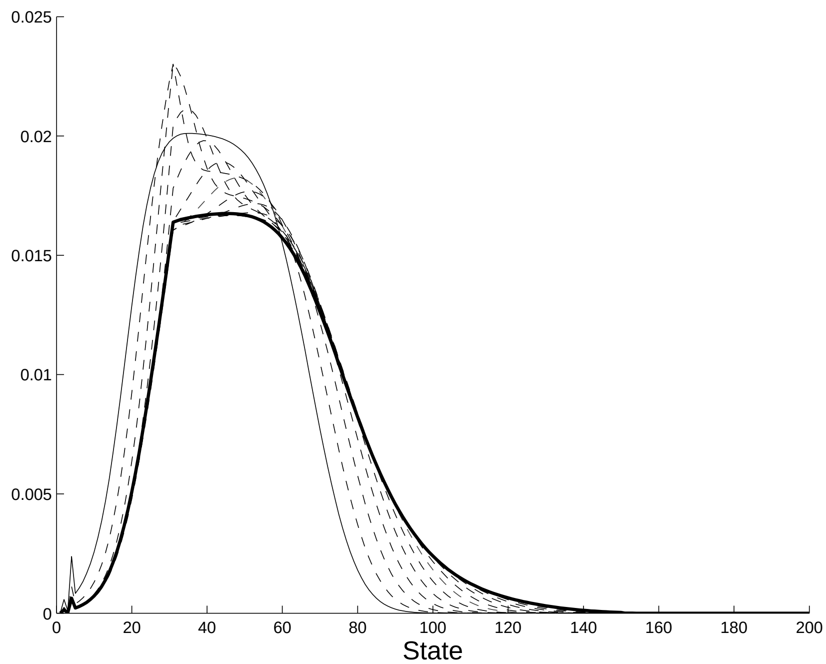

Figure 5.10: Health state distribution of active plan participants, $\underline{p}_t^{(A)}$
$x_l = 10$, $x_u = 30$, SW1911M.

where λ_A is the maximal eigenvalue of the matrix Π_{AA}.

Consider the *average aggregate seniority* $\hat{r}_t$ given by Eq. (4.27). Due to the generator property of the matrix Π, all the quantities in the model converge, so does $\hat{r}_t$. This we illustrate in Fig. 5.11, in which $\hat{r}_t$ converges to 7 at about 13 years.

We examine the stationary health state distribution of plan participants in a pension system, perturbed at time τ_m with respect to the mortality rates, as shown in Eq. (5.2). Denote by $\tilde{\Pi}$ the perturbed generator, and by $\underline{\tilde{\pi}}$ the stationary distribution of states. One method to obtain $\underline{\tilde{\pi}}$ is to find it directly from the system given in Eq. (1.19)

$$\begin{cases} \underline{\tilde{\pi}} e^{\tilde{\Pi}} = \underline{\tilde{\pi}}, \\ \underline{\tilde{\pi}} \mathbf{1} = 1. \end{cases} \tag{5.22}$$

Another method is to recall the results of the Markov chains perturbation analysis described in G. E. Cho and C. D. Meyer [19] and E. Seneta [55]. In the current framework we consider one generic type of the perturbation and a special example.

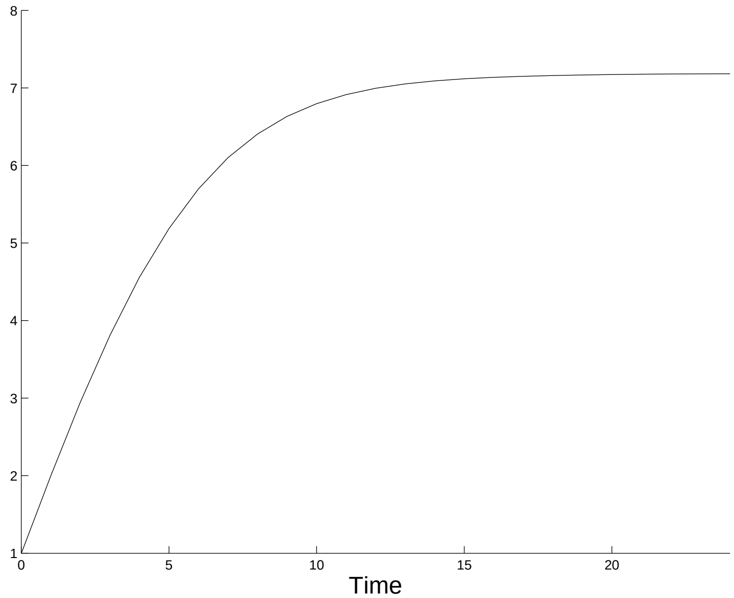

Figure 5.11: Average aggregate seniority, $\hat{r}_t$, SW1911M

Generic Perturbation. We consider the same perturbation as in Eq. (5.2), where

$$\tilde{\Pi} = \Pi + \varepsilon \hat{D}(\Pi_{AD}),$$

where ε is a positive scalar and $\hat{D}(\Pi_{AD})$ is defined by Eq. (5.3). According to G. E. Cho and C. D. Meyer [19], proposition 2.1, the difference between the stationary distribution of the original and the perturbed system is given by

$$\underline{\tilde{\pi}} - \underline{\pi} = \underline{\tilde{\pi}} \varepsilon \hat{D}(\Pi_{AD})(-\Pi)^{\#}, \tag{5.23}$$

and where $(-\Pi)^{\#}$ is the *group inverse* of $(-\Pi)$.

Let A be a square matrix and denote by $A^{\#}$ its *group inverse*. According to G. E. Cho and C. D. Meyer [19], $A^{\#}$ is the unique matrix that satisfies

- $AA^{\#}A = A$;
- $A^{\#}AA^{\#} = A^{\#}$;
- $AA^{\#} = A^{\#}A$.

We define $A = (-\Pi)$ and make the following partition of matrix A and $\underline{\pi}$

$$A = \begin{pmatrix} A_N & \underline{c} \\ \underline{d}^T & A_{(N,N)} \end{pmatrix}, \quad \underline{\pi} = (\underline{\bar{\pi}}, \, \pi_N),$$

where N is a chosen state. The group inverse of A is given by

$$A^{\#} = \begin{pmatrix} (I - \mathbf{1}\underline{\bar{\pi}})A_N^{-1}(I - \mathbf{1}\underline{\bar{\pi}}) & -\pi_N(I - \mathbf{1}\underline{\bar{\pi}})A_N^{-1}\mathbf{1} \\ -\underline{\bar{\pi}}A_N^{-1}(I - \mathbf{1}\underline{\bar{\pi}}) & \pi_N\underline{\bar{\pi}}A_N^{-1}\mathbf{1} \end{pmatrix}.$$

The proposition also defines the difference between π_N and $\tilde{\pi}_N$

$$\tilde{\pi}_N - \pi_N = \tilde{\pi}_N \varepsilon \hat{D}(\Pi_{AD})A^{\#}_{*N},$$

where $A^{\#}_{*N}$ denotes the N^{th} column of $A^{\#}$.

In Fig. 5.12 we present the stationary distributions of active health states, obtained from Eq. (5.22) for the original and the perturbed system. In Fig. 5.13 the solid line represents the difference between $\tilde{\underline{\pi}}$ and $\underline{\pi}$, obtained from Eq. (5.23), or direct computation, the dashed line – from the method given by Eq. (5.22). In this example, we chose 10^{-3} as the degree of precision for the stability time calculation. Both stationary probabilities are zero for the states greater than 150, which corresponds to the retirement age 65. Our general expected observation is that in the perturbed pension system there are less active participants in young health states, and more active participants in older health states. Also, the decrease of the mortality rates leads to the increase of the time spent among active states, which, evidently, increases the total number of active participants in the perturbed pension system. This theoretical fact we confirm by performing the numerical summation of the differences between the probabilities, given by the graphs.

Special example. Let us consider a special case, where we change only two elements of generator Π

$$\tilde{\Pi}_{(i^*,i^*)} = \Pi_{(i^*,i^*)} - r, \quad \tilde{\Pi}_{(i^*,R)} = \Pi_{(i^*,R)} + r,$$

where, i^* is a chosen health state, r is a positive constant and R is one of the inactive states, for example, "retirement". We reorganize the i^*-th row by moving some probability mass to the phase R . This type of perturbation is considered in Corollary 4.2 in G. E. Cho and C. D. Meyer [19]. According to the corollary, the relative change in the stationary probability of the perturbed state i^* is given by

$$\pi_{i^*} - \tilde{\pi}_{i^*} = r\tilde{\pi}_{i^*}(A^{\#}_{(i^*,i^*)} - A^{\#}_{(i^*,R)}).$$

The corollary also shows that the closer the distance between states i^* and R, the less sensitive π_{i^*} to the perturbation.

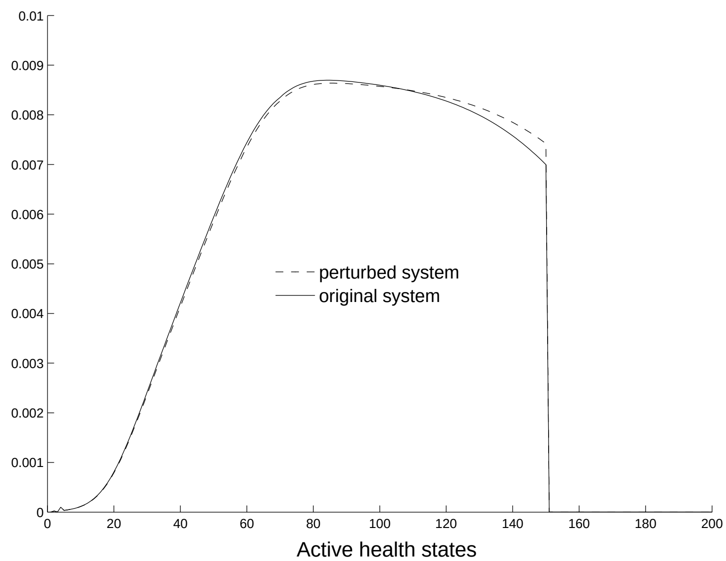

Figure 5.12: Stationary health state distributions, $\underline{\tilde{\pi}}$ and $\underline{\pi}$ $\varepsilon = 0.3$, SW1911M

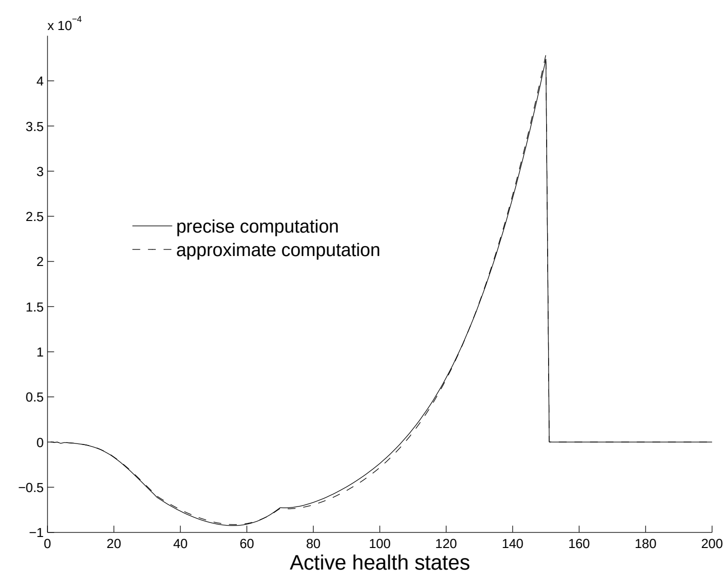

Figure 5.13: Perturbation analysis, $\underline{\tilde{\pi}} - \underline{\pi}$, $\varepsilon = 0.3$, SW1911M

Chapter 6

Profits&Losses: post-retirement

In this chapter we construct a profit-test model for the post-retirement period with the main purpose to examine the impact of health on the present value of future cash flows. As we have discussed in the introduction to Part II, the plan participants evolve in time in accordance with the PH-aging process introduced in Sec. 2.1. Newly retired pensioners have health states distribution given by $\underline{\tau}_R$ defined by Eq. (2.3).

In Section 6.1 we obtain the distribution of the present value of the cash flows assuming a deterministic number of new arrivals each year, and in Section 6.2 we consider a Poisson process for the arrivals. We compare the two approaches in the end of Section 6.2. For both approaches we examine in Section 6.3 the impact of health on the distribution of the present value.

6.1 Deterministic arrival of new participants

We assume here that a constant number N of new pensioners joins the fund at the beginning of each year. We define by $\psi_{u,t}$ the present value at time u of the cash flow over the interval $[u, t)$ for a pensioner who joined the fund at the beginning of year u, $0 \leq u < t$. The total present value at time 0 for the fund over the interval

$[0, t)$ is

$$V_t = \sum_{u=0}^{t-1} v^u \sum_{i=1}^{N} \psi_{u,t}^{(i)}, \qquad \text{for } t = 0,\, 1,\, \ldots \tag{6.1}$$

where $\psi_{u,t}^{(i)}$, $i = 1, \ldots, N$ are i.i.d. random variables and they have the same distribution as $\psi_{u,t}$. Here, we apply the discount coefficient v^u to $\psi_{u,t}$ in order to obtain the present value at time 0. To determine the distribution of $\psi_{u,t}$, we condition on the number of years spent as a pensioner, and obtain the density displayed in Table 6.1, where $\underline{\tau}_R$ and Λ are defined by Eq. (2.3) and Eq. (2.2), respectively.

Event	Value	Probability
still alive after $t - u$ years	$\ddot{a}_R - \sum_{k=0}^{t-u-1} v^k$	$\underline{\tau}_R e^{\Lambda(t-u)}\mathbf{1}$
dies in $[r, r+1)$, $0 \le r \le t-u-1$	$\ddot{a}_R - \sum_{k=0}^{r} v^k$	$\underline{\tau}_R e^{\Lambda r}(\mathbf{1} - e^{\Lambda}\mathbf{1})$

Table 6.1: Distribution of $\psi_{u,t}$.

We observe from Table 6.1 that the distribution of $\psi_{u,t}$ depends only on the difference $t-u$, and we denote by μ_h and θ_h^2, respectively, the mean and the variance of $\psi_{u,u+h}$, where h is the length of the interval spent by the pensioner in the pension plan.

Theorem 6.1.1 *The mean μ_h and the variance θ_h^2 are given by*

$$\mu_h = \ddot{a}_R - \ddot{a}_{R:h} \tag{6.2}$$

$$= v^h \underline{\tau}_R e^{\Lambda h}(I - ve^{\Lambda})^{-1}\mathbf{1}, \tag{6.3}$$

$$\theta_h^2 = \underline{\tau}_R \sum_{k=0}^{h-1} v^k e^{\Lambda k}\mathbf{1}\gamma_k - (\ddot{a}_{R:h})^2, \tag{6.4}$$

where $\ddot{a}_R$ and $\ddot{a}_{R:h}$ are given by Eq. (2.15), (2.16), *and*

$$\gamma_k = \frac{2 - v^k(1+v)}{1-v}, \qquad \textit{for } k = 0,\, 1,\, \ldots \tag{6.5}$$

Proof. Eq. (6.2) is a classical result, which may be obtained as follows from the probability distribution in Table 6.1.

$$\begin{aligned}\mu_h &= \ddot{a}_R - \sum_{r=0}^{h-1} \underline{\tau}_R e^{\Lambda r}(\mathbf{1} - e^{\Lambda}\mathbf{1}) \sum_{k=0}^{r} v^k - \underline{\tau}_R e^{\Lambda h}\mathbf{1} \sum_{k=0}^{h-1} v^k \\ &= \ddot{a}_R - \sum_{k=0}^{h-1} v^k \underline{\tau}_R e^{\Lambda k}\mathbf{1} \qquad \text{after rearranging the sums} \\ &= \ddot{a}_R - \sum_{k=0}^{h-1} v^k S_R(k) \ = \ddot{a}_R - \ddot{a}_{R:h},\end{aligned}$$

where $S_R(k)$ is defined in Eq. (2.4). The expression (6.3) follows directly from Eq. (2.15) and Eq. (2.16).

To prove Eq. (6.4) we start from

$$E[\psi^2_{u,u+h}] = (\ddot{a}_R - \sum_{k=0}^{h-1} v^k)^2 \underline{\tau}_R e^{\Lambda h}\mathbf{1} + \sum_{r=0}^{h-1} (\ddot{a}_R - \sum_{k=0}^{r} v^k)^2 \underline{\tau}_R e^{\Lambda r}(\mathbf{1} - e^{\Lambda}\mathbf{1}),$$

and we reorganize the sums in increasing power of e^{Λ}. After simple, but tedious calculations, we obtain

$$E[\psi^2_{u,u+h}] = \ddot{a}_R^2 - 2\ddot{a}_R\ddot{a}_{R:h} + \underline{\tau}_R \sum_{k=0}^{h-1} e^{\Lambda k}\mathbf{1} v^k \gamma_k, \tag{6.6}$$

and, using the formula $\theta_h^2 = E[\psi^2_{u,u+h}] - \mu_h^2$, we find that

$$\theta_h^2 = \underline{\tau}_R \sum_{k=0}^{h-1} e^{\Lambda k}\mathbf{1} v^k \gamma_k - (\ddot{a}_{R:h})^2.$$

■

Pensioners are assumed to evolve independently in a fixed mortality environment. Thus, by the central limit theorem, each partial sum in Eq. (6.1) may be approximated by a normal random variable, for N large enough, and we conclude that V_t is a sum of discounted normal random variables that are independent. Therefore, V_t has a normal distribution with parameters given by

$$E[V_t] = N \sum_{s=0}^{t-1} \mu_{t-s} v^s, \qquad Var[V_t] = N \sum_{s=0}^{t-1} \theta^2_{t-s} v^{2s}. \tag{6.7}$$

We observe that

$$\lim_{t\to\infty} E[V_t] = 0. \tag{6.8}$$

To show this, we use (6.3) and write

$$\begin{aligned} E[V_t] &= N\underline{\tau}_R \sum_{s=0}^{t-1} v^{t-s} e^{\Lambda(t-s)} (I - ve^{\Lambda})^{-1} \mathbf{1} v^s \\ &= v^t N\underline{\tau}_R \sum_{s=0}^{t-1} e^{\Lambda(t-s)} (I - ve^{\Lambda})^{-1} \mathbf{1} \\ &= v^t N\underline{\tau}_R e^{\Lambda}(I - e^{\Lambda t})(I - e^{\Lambda})^{-1}(I - ve^{\Lambda})^{-1}\mathbf{1}. \end{aligned} \tag{6.9}$$

To understand for which values of t it is reasonable to apply the model, we determine time t when $E[V_t]$ becomes approximately zero. We fix a small ϵ as a required degree of precision and look for t^* such that $|E[V_t]| < \epsilon$ for $t > t^*$. The only time-dependent component in Eq. (6.9) is the product $v^t(I - e^{\Lambda t})$, so that the problem can be reformulated as follows. We look for t^* such that $\omega(t) = v^t(1 - \lambda_*^t) < \epsilon$ for all $t > t^*$ where λ_*, $\lambda_* < 1$ is the dominant eigenvalue of e^{Λ}. Thus, t^* is determined by $\omega(t) = \epsilon$. In Fig. 6.1 we plot $E[V_t]$ and we indicate t^* computed for different values of ϵ. One sees from the figure that the impact of ϵ is quite significant: t^* is about 48 years for $\epsilon = 10^{-2}$, 73 years for $\epsilon = 10^{-3}$ and 97 years for $\epsilon = 10^{-4}$. Actually, the shape of V_t as a function of t is explained by the shape of $\omega(t)$, which we display in Fig. 6.2.

The variance ${\theta_h}^2$ converges to the positive constant

$$\theta_\infty^2 = \lim_{h\to\infty} \theta_h^2 = \underline{\tau}_R \sum_{k=0}^{\infty} v^k e^{\Lambda k} \mathbf{1} \gamma_k - (\ddot{a}_R)^2.$$

To deal with $Var[V_t]$ as $t \to \infty$, we need the following technical lemma.

Lemma 6.1.2 *Assume that* $f(t)$, $t = 1, 2, ...$ *is a positive finite function and such that* $\lim_{t\to\infty} f(t) = f_\infty < \infty$, *and assume that* $0 < \nu < 1$. *One has*

$$\lim_{t\to\infty} \sum_{s=0}^{t-1} f(t-s)\nu^s = \frac{f_\infty}{1-\nu}. \tag{6.10}$$

Proof. Define $f(s) = 0$ for $s \leq 0$ and $g(t) = \sum\limits_{s=0}^{\infty} f(t-s)\nu^s$. Then,

$$\lim_{t\to\infty} g(t) = \lim_{t\to\infty} \sum_{s=0}^{\infty} f(t-s)\nu^s,$$

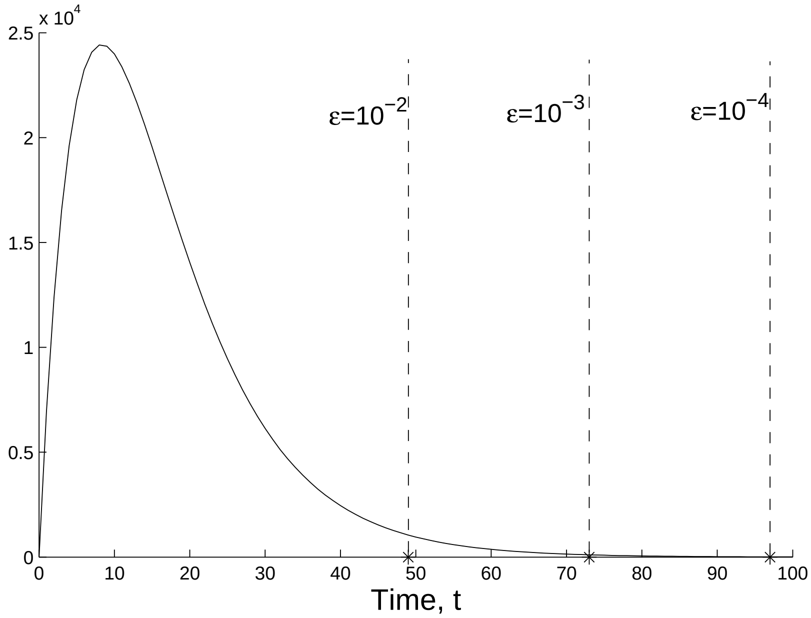

Figure 6.1: Expected present value, $E[V_t]$
Parameters: SW1911M, $v = 0.9091$, $R = 65$, $N = 1000$

and, by the dominated convergence theorem,

$$\begin{aligned}\lim_{t\to\infty} g(t) &= \sum_{s=0}^{\infty}(\lim_{t\to\infty} f(t-s))\nu^s \\ &= \sum_{s=0}^{\infty} f_\infty \nu^s = f_\infty \sum_{s=0}^{\infty} \nu^s \\ &= \frac{f_\infty}{1-\nu}.\end{aligned}$$

■

The convergence of $Var[V_t]$, given by Eq. (6.7), immediately results from Lemma 6.1.2 with $f(t) = \theta_t^2$, $f_\infty = \theta_\infty^2$, $\nu = v^2$ and

$$\lim_{t\to\infty} Var[V_t] = \frac{N\theta_\infty^2}{1-v^2}.$$

We illustrate $Var[V_t]$ in Fig. 6.3.

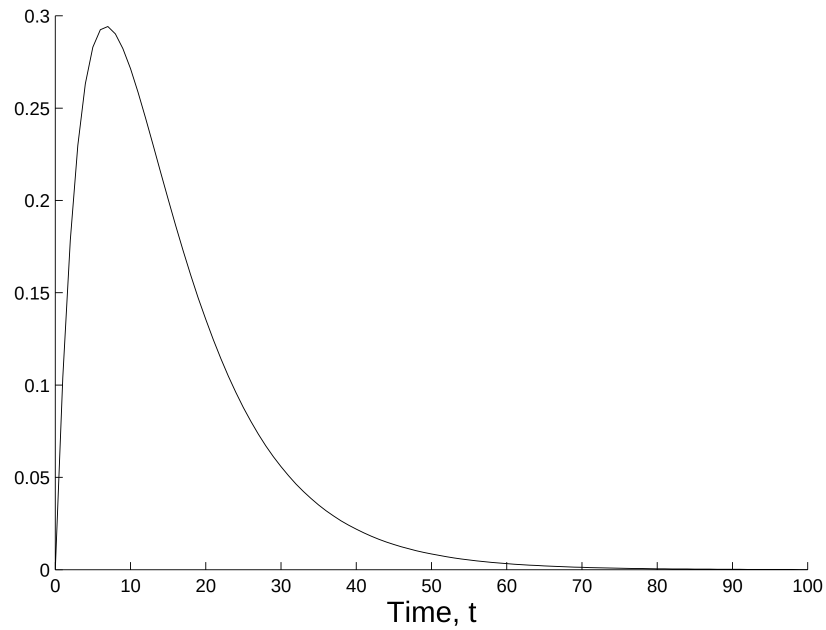

Figure 6.2: $\omega(t)$. Parameters: SW1911M, $v = 0.9091$

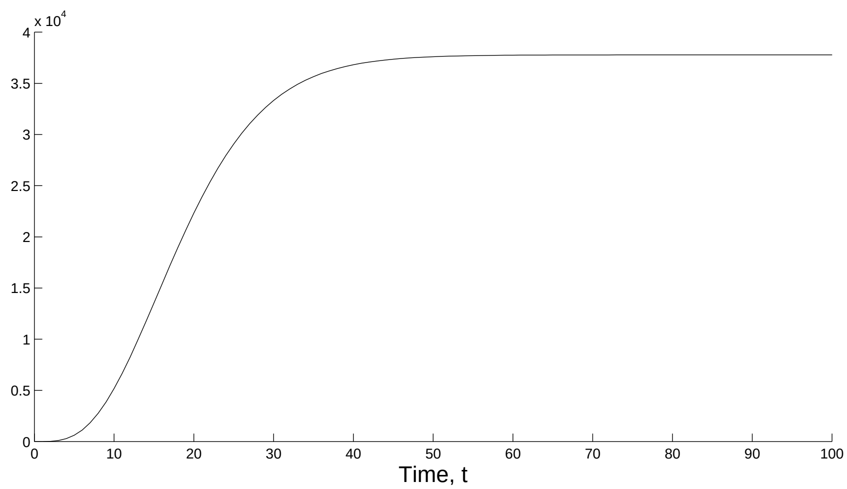

Figure 6.3: Variance $Var[V_t]$
Parameters: SW1911M, $v = 0.9091$, $R = 65$, $N = 1000$

6.2 Poisson process for new participants

In this section, time is continuous and we assume that new pensioners arrive according to a Poisson process. With this assumption, the number of new pensioners N_u in year $[u, u+1)$, $0 \leq u \leq t-1$ is a Poisson random variable with parameter N. We denote by K_t the total number of new pensioners during the interval $[0, t)$ and by W_t the total present value at time 0 of the cash flow over the interval $[0, t)$. Then,

$$W_t = \sum_{u=1}^{t-1} v^u \sum_{i=1}^{N_u} \psi_{u,t}^{(i)}, \qquad \psi_{u,t}^{(i)} \sim \psi_{u,t}. \tag{6.11}$$

Conditionally given that $N_0 + ... + N_{t-1} = K$, the years of arrival have the same distribution as K iid random variables uniform in $\{0, 1, ..., t-1\}$. Thus,

$$W_t \stackrel{d}{=} \sum_{i=1}^{K_t} \phi_t^{(i)}, \tag{6.12}$$

where $\phi_t^{(i)}$ are independent and identically distributed random variables distributed as ϕ_t. The distribution of ϕ_t is the same as of $v^u \psi_{u,t}$ with u being a uniform random variable in $\{0, 1, ..., t-1\}$. The total number K_t has a Poisson distribution with parameter Nt, so W_t is a compound Poisson random variable.

Theorem 6.2.1 *The mean M_t and the variance Θ_t^2 of ϕ_t, $t \geq 1$, integer, are*

$$\begin{aligned} M_t &= \frac{1}{t} \sum_{k=0}^{t-1} \mu_{t-k} v^k &(6.13)\\ &= \frac{v^t}{t} \underline{\mathcal{I}}_R e^{\Lambda} (I - e^{\Lambda t})(I - e^{\Lambda})^{-1} (I - v e^{\Lambda})^{-1} \mathbf{1}, &(6.14)\\ \Theta_t^2 &= \frac{1}{t} \sum_{k=0}^{t-1} \theta_{t-k}^2 v^{2k} + \frac{1}{t} \sum_{k=0}^{t-1} \mu_{t-k}^2 v^{2k} - \left(\frac{1}{t} \sum_{k=0}^{t-1} \mu_{t-k} v^k \right)^2. &(6.15) \end{aligned}$$

Proof. As we have indicated above, ϕ_t is a uniform mixture of the random variables $v^u\psi_{u,t}$, for $0 \leq u \leq t-1$. Therefore,

$$\begin{aligned}
E[\phi_t] &= \frac{1}{t}\sum_{k=0}^{t-1} E[\psi_{k,t}]v^k = \frac{1}{t}\sum_{k=0}^{t-1}\mu_{t-k}v^k \\
&= \frac{1}{t}\sum_{k=0}^{t-1} v^t \underline{\mathcal{I}}_R e^{\Lambda(t-k)}(I - ve^{\Lambda})^{-1}\mathbf{1} \qquad \text{by (6.3)} \\
&= \frac{v^t}{t}\underline{\mathcal{I}}_R \sum_{k=0}^{t-1} e^{\Lambda(t-k)}(I - ve^{\Lambda})^{-1}\mathbf{1} \\
&= \frac{v^t}{t}\underline{\mathcal{I}}_R (I - e^{\Lambda t})(I - e^{\Lambda})^{-1}(I - ve^{\Lambda})^{-1}\mathbf{1}.
\end{aligned}$$

For the second moments, we have

$$E[\phi_t^2] = \frac{1}{t}\sum_{k=0}^{t-1} E[\psi_{k,t}^2]v^{2k},$$

so that

$$\begin{aligned}
\Theta_t^2 &= \frac{1}{t}\sum_{k=0}^{t-1} E[\psi_{k,t}^2]v^{2k} - \left(\frac{1}{t}\sum_{k=0}^{t-1} E[\psi_{k,t}]v^k\right)^2 \\
&= \frac{1}{t}\sum_{k=0}^{t-1}\theta_{t-k}^2 v^{2k} + \frac{1}{t}\sum_{k=0}^{t-1}\mu_{t-k}^2 v^{2k} - \left(\frac{1}{t}\sum_{k=0}^{t-1}\mu_{t-k}v^k\right)^2.
\end{aligned}$$

The three terms represent the average variance, the average squared mean and the average mean squared of $v^k\psi_{k,t}$, respectively. ■

Since $E[W_t] = NtM_t$, it is immediately obvious from (6.9, 6.14) that $E[W_t] = E[V_t]$, so that the two models give the same expected present values. The variance of W_t can be expressed in terms of the variance of V_t as follows

$$Var[W_t] = NtE[\phi_t^2] \;= N\sum_{k=0}^{t-1} E[\psi_{k,t}^2]v^{2k} \tag{6.16}$$

$$= N\sum_{k=0}^{t-1}(\theta_{t-k}^2 + \mu_{t-k}^2)v^{2k} \;= Var[V_t] + N\sum_{k=0}^{t-1}\mu_{t-k}^2 v^{2k}, \tag{6.17}$$

which shows that:

- $Var[W_t]$ is greater than or equal to $Var[V_t]$;
- $Var[W_t]$ and $Var[V_t]$ converge to the same constant as $t \to \infty$: we apply Lemma 6.1.2 with $f(t) = \mu_t^2$, $f_\infty = \lim_{t\to\infty} \mu_t = 0$ and $\nu = v^2$ and find that the right-most term in (6.17) converges to zero.

We illustrate the indicated relations between $Var[W_t]$ and $Var[V_t]$ in Fig. 6.4. The dashed line stands for $Var[W_t]$ and the solid line stands for $Var[V_t]$. One observes from the figure that the shape of $Var[W_t]$ is similar to the shape of $\omega(t)$ presented in Fig. (6.2). This is because $N \sum_{k=0}^{t-1} \mu_{t-k}^2 v^{2k}$ has a similar behavior as $\omega(t)$.

We conclude that the model with Poisson arrivals may be chosen for short-term computations as it allows us to capture the risks related to the uncertainty in the arrivals. In the long-term, the effect of this uncertainty becomes less significant, and therefore one may use the model with deterministic arrivals.

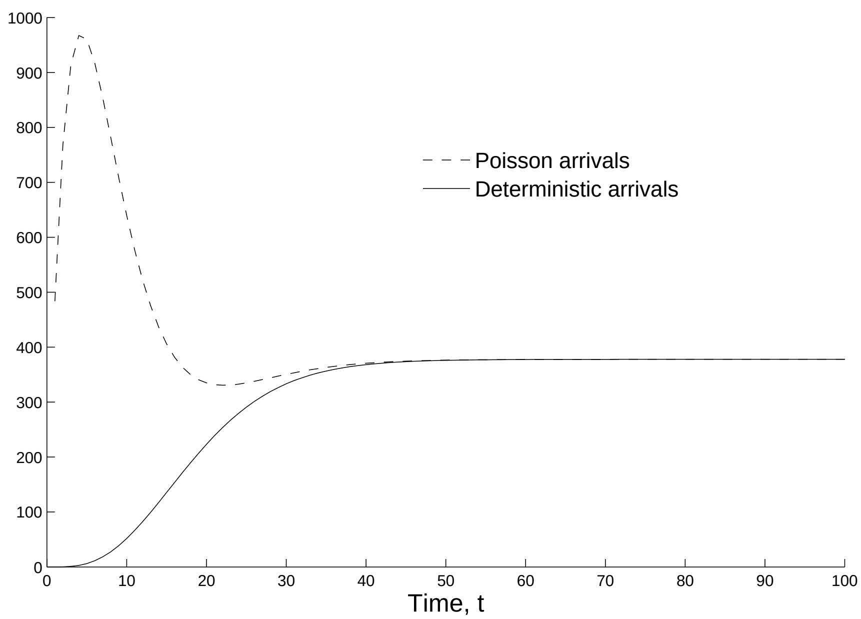

Figure 6.4: Variances $Var[W_t]$ and $Var[V_t]$
Parameters: SW1911M, $N = 10$, $R = 65$, $v = 0.9091$

Computational aspects. The iterative *Panjer recursion*, introduced in Panjer [49] and presented in Appendix A.5, is traditionally used to find the density function of a positive compound random variable of the *Panjer class*. The present value W_t

is a compound Poisson random variable consisting of individual present values ϕ_t. Individual present values ϕ_t may take negative values when the life of an individual is longer than expected and, therefore, W_t may also be negative. In order to compute the density function of W_t, we follow the procedure suggested in Sundt and Jewell [58]. We introduce two random variables:

$$\phi_t^+ = \max(0, \phi_t), \qquad \text{and} \qquad \phi_t^- = \max(0, -\phi_t), \tag{6.18}$$

and also define

$$W_t^+ = \sum_{s=1}^{N(t)} \phi_s^+, \qquad \text{and} \qquad W_t^- = \sum_{s=1}^{N(t)} \phi_s^-,$$

then $W_t = W_t^+ - W_t^-$. Thus, one needs to apply the Panjer recursion twice to obtain the distribution of W_t^- and W_t^+, and finally to take the *convolution* of W_t^+ and $(-W_t^-)$, defined in Appendix A.4, to obtain the distribution of W_t. Note that W_t^+ and W_t^- are independent because $N(t)$ has the Poisson distribution and the ϕ_is are independent.

A question one may ask is whether it is worth the effort of computing the exact distribution when we may as well use the normal approximation based on the first two moments. We compare in Fig. 6.5 the density of W_t obtained by Panjer's algorithm and the normal density function. We observe that even for small values of N the approximation is very good. The dashed line is the density of W_t obtained from the extended Panjer procedure and the solid line is the normal approximation with the same mean and variance. It is easy to see that the two distributions are almost identical.

A second problem with the Panjer recursion appears when the parameter of the Poisson distribution of K_t is big. That parameter is Nt, and, according to Eq. (A.13), the iterations in the Panjer recursion start from $\Pr[\,K_t = 0\,] = e^{-Nt}$, this may be identical to zero up to machine precision, and may result in all probabilities being miscomputed as zero. In order to overcome this difficulty, we have started with an arbitrary value at the first iteration and re-normalize the probabilities at the end of each iteration. In this case, we have stopped the algorithm when the mean and the variance from the calculated distribution are close to the theoretical values from Eq. (6.9, 6.16).

In Fig. 6.5 we also display the normal approximation for the density of V_t as a dotted line. As we have discussed, the expected values of W_t and V_t are identical, and the variances differ as shown in Eq. (6.17) and Fig. 6.4. In Fig. 6.5 we choose $t = 15$ years, which corresponds to a moderate difference between the variances.

In some applications it is important to calculate the distribution of W_t up to extremely high quantiles. One such example is the loss of credit portfolios of a

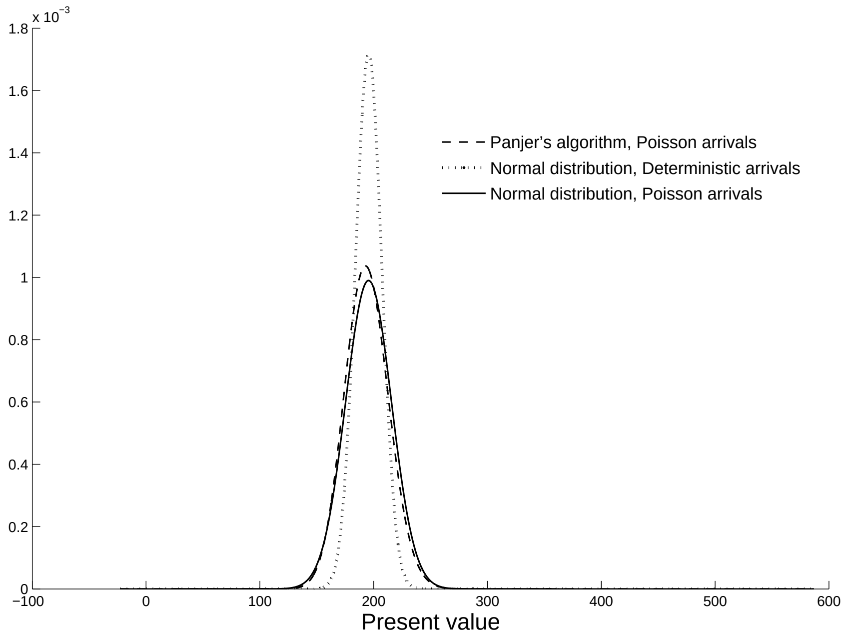

Figure 6.5: Probability density function of the total present value
Parameters: SW1911M, $N = 10$, $t = 15$, $v = 0.9091$

bank, which in accordance with Basel II regulations has to be calculated up to the 99% quantile level. In S. Gerhold et al. [27] the authors derive numerically stable algorithms based on iterative Panjer's recursion, which allow to make precise calculations.

6.3 Impact of health

We investigate in this section the impact on W_t and V_t of a general change of health of the pensioners. As we have mentioned in Section 2.4, an unexpected change of health conditions causes additional uncertainties about the remaining lifetime of individuals and is called a longevity risk. In Section 2.4 we identified two different major causes of longevity. The first is related to internal factors, that is, properties of the human body which are defined by genetics and by personal habits. The second is related to external factors that externally affects the life of individual, like economics, medical service, scientific developments, etc. We now examine the impact of each cause on the present values W_t and V_t.

Internal factors. As we suggested in Section 2.4, one may assume that the individuals retiring at age R have the health distribution given by $\underline{\tau}_R^*$. According to Eq. (2.28), the new survival function is $S_R^*(t) = \underline{\tau}_R^* e^{\Lambda t}\mathbf{1}$.

We assume that the annuities $\ddot{a}_R$ are computed with initial health distribution $\underline{\tau}_R$, but in actual fact retires with $\underline{\tau}_R^*$. Therefore, to analyze the impact of health we change only the probabilities $\underline{\tau}_R$ by $\underline{\tau}_R^*$ in Table 6.1 to obtain the distribution of new random variable denoted by $\psi_{u,u+h}^*$; everything else remains the same. The total present value V_t^* over the interval $[0, t)$ is

$$V_t^* = \sum_{u=0}^{t-1} v^u \sum_{i=1}^{N} {\psi_{u,t}^*}^{(i)}, \qquad \text{for } t = 0,\, 1,\, \ldots$$

Let us denote by μ_h^* and θ_h^{*2}, respectively, the mean and the variance of $\psi_{u,u+h}^*$. The expressions for μ_h^* and θ_h^{*2} are given in the following lemma.

Lemma 6.3.1 *Assume that the random variable $\psi_{u,u+h},\ t \geq 0$ is defined by Table 6.1, where the health state distribution is $\underline{\tau}_R^*$. Its mean and variance are given by*

$$\begin{aligned} \mu_h^* &= (\underline{\tau}_R - \underline{\tau}_R^* + v^h \underline{\tau}_R^* e^{\Lambda h})(I - ve^{\Lambda})^{-1}\mathbf{1}, & (6.19)\\ \theta_h^{*2} &= \underline{\tau}_R^* \sum_{k=0}^{h-1} v^k e^{\Lambda k}\mathbf{1}\gamma_k - (\ddot{a}_{R:h}^*)^2, & (6.20)\end{aligned}$$

where $\ddot{a}_{R:h}^* = \underline{\tau}_R^*(I - v^h e^{\Lambda h})(I - ve^{\Lambda})^{-1}\mathbf{1}$ and the γ_k's are given in Eq. (6.5). *In particular, if $\underline{\tau}_R^* = \underline{\tau}_{R-\gamma},\ 0 \leq \gamma \leq R$, then*

$$\begin{aligned} \mu_h^* &= \ddot{a}_R - \ddot{a}_{R-\gamma:h}, & (6.21)\\ \theta_h^{*2} &= \underline{\tau}_R^* \sum_{k=0}^{h-1} v^k e^{\Lambda k}\mathbf{1}\gamma_k - (\ddot{a}_{R-\gamma:h})^2. & (6.22)\end{aligned}$$

Proof. The proof repeats the one of Theorem 6.1.1. However, when $\underline{\tau}_R^* = \underline{\tau}_{R-\gamma},\ 0 \leq \gamma \leq R$ it is simply enough to replace R by $R - \gamma$ in Eq. (6.2)-(6.4). ∎

For the model with Poisson arrivals, the expectation and the variance of ϕ_t^* can be found using their relation to μ_h^* and θ_h^{*2} defined by Theorem 6.2.1. Specifically, for M_t^* and Θ_t^{*2} we have the following lemma.

Lemma 6.3.2 *The mean M_t^* and the variance Θ_t^{*2} of ϕ_t^*, $t \geq 1$, with the health state distribution $\underline{\tau}_R^*$ at retirement, are*

$$
\begin{aligned}
M_t^* &= \frac{1}{t}\sum_{k=0}^{t-1} \mu_{t-k}^* v^k \\
\Theta_t^{*2} &= \frac{1}{t}\sum_{k=0}^{t-1} (\theta_{t-k}^*)^2 v^{2k} + \frac{1}{t}\sum_{k=0}^{t-1} (\mu_{t-k}^*)^2 v^{2k} - \left(\frac{1}{t}\sum_{k=0}^{t-1} \mu_{t-k}^* v^k\right)^2,
\end{aligned}
$$

where μ_{t-k}^* and $(\theta_{t-k}^*)^2$ are given by Eq. (6.19) *and Eq.* (6.20), *respectively.*

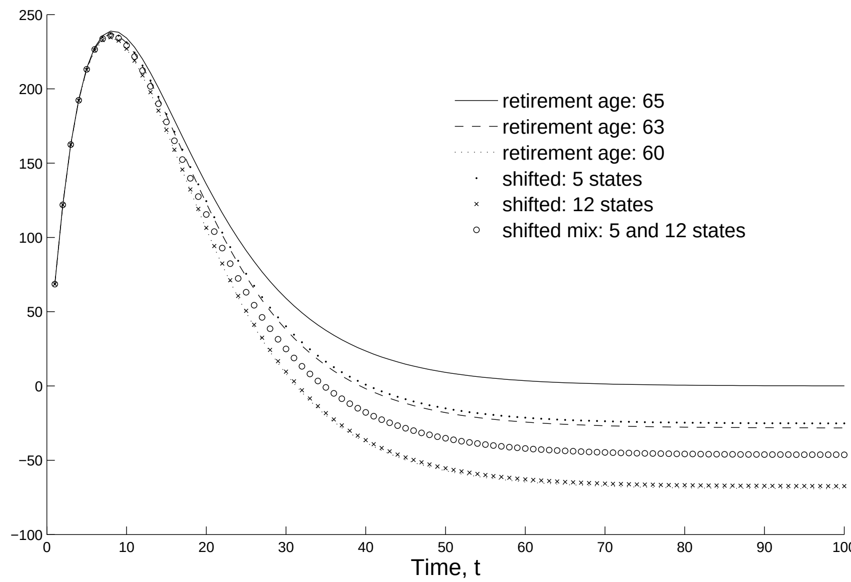

Figure 6.6: Impact of internal factors: $E[V_t^*]$.
Parameters: SW1911M, $N = 10$, $v = 0.9091$

We illustrate the impact of internal factors on the expected present value $E[V_t^*]$ in Fig. 6.6. We show the expectation for different retirement ages, 60, 63 and 65 years old with $\gamma = 5$, 2 and 0, respectively. As $\ddot{a}_R$ is computed for $\underline{\tau}_{65}$, the expectation for early retirements at age 63 and at age 60 converges to negative constants. Obviously, the earlier the retirement, the better the health, the longer the life, and, therefore, the lower the present value.

We also recall from Section 2.4 that we may shift the distribution of health states at retirement. In this particular example, we see that to decrease the retirement age to 60 has the same impact on the expectation as shifting the distribution of health states at age 65 by $\delta = 12$ phases: the corresponding curves are nearly identical. Similarly, to decrease the retirement age to 63 is the same as to shift the distribution of health states by $\delta = 5$ phases. The expectations are shown in different types of dots.

It must be noted that the health state distribution may be changed in other manners. For instance, define by $\underline{\tau}_1^*$ the distribution of health states shifted by 5 states with respect to $\underline{\tau}_R$; by $\underline{\tau}_2^*$ the distribution of health states shifted by 12 states with respect to $\underline{\tau}_R$. Then, the choice of $\underline{\tau}^* = 1/2\underline{\tau}_1^* + 1/2\underline{\tau}_2^*$ corresponds to the situation where we have two populations of mixed health. The expectation is shown as circles.

External factors. Here, we assume that at time K the mortality rates become lower for all participants. We apply the technique described in Section 2.4 to obtain the distribution of the perturbed individual present value $\psi^{\varepsilon}_{0,t}(K)$ given in Table 6.2 with ε being the parameter of the perturbation. The difference with Table 6.1 is in the column of probabilities. Here, we divide the life of the participant into two periods: before and after the change of mortality rates. The present values in Table 6.2 remain the same as in Table 6.1, because the annuity $\ddot{a}_R$ is calculated before the change of mortality rates. The total perturbed present value V_t^{ε} over the interval $[0, t)$ is

$$V_t^{\varepsilon} = \sum_{u=0}^{t-1} v^u \sum_{i=1}^{N} \psi^{\varepsilon}_{u,t}, \qquad \text{for } t = 0,\, 1,\, \ldots$$

Obviously, for $K > t$ the expectation and the variance of $\psi^{\varepsilon}_{0,t}(K)$ are the same as of $\psi_{0,t}$ and is defined by Theorem 6.1.1. The distributions of $\psi^{\varepsilon}_{u,t}(K)$, $0 < u \le t-1$ and $K \le t$ can be obtained on the basis of the distribution $\psi^{\varepsilon}_{0,t}(K)$ as follows

$$\begin{aligned} \psi^{\varepsilon}_{u,t}(K) &\overset{d}{=} \psi^{\varepsilon}_{0,t-u}(K-u), && 1 \le u \le K, \\ \psi^{\varepsilon}_{u,t}(K) &\overset{d}{=} \psi^{\varepsilon}_{0,t-u}(0), && K+1 \le u \le t-1. \end{aligned}$$

For $K \le t$, the expectation and the variance of $\psi^{\varepsilon}_{0,t}(K)$ are given in the following Lemma.

Lemma 6.3.3 *Assume that the distribution of $\psi^{\varepsilon}_{0,t}(K)$, $K \le t$ is defined by Table 6.2. The mean μ_t^{ε} and the variance $\theta_t^{\varepsilon 2}$ of $\psi^{\varepsilon}_{0,t}(K)$ are given by*

$$\mu_t^{\varepsilon} = \mu_K - S_R(K) v^K (\ddot{a}^{\varepsilon}_{R+K:t-K-1} - 1), \tag{6.23}$$
$$\theta_t^{\varepsilon 2} = \theta^2_{K+1} + S_R(K) v^K \Delta^{\varepsilon}_{t,K}, \tag{6.24}$$

Event	Value	Probability
still alive after t years, $t \le K$	$\ddot{a}_R - \sum_{k=0}^{t-1} v^k$	$\underline{\tau}_R e^{\Lambda t}\mathbf{1}$
still alive after t years, $t > K$	$\ddot{a}_R - \sum_{k=0}^{t-1} v^k$	$\underline{\tau}_R e^{\Lambda K} e^{\tilde{\Lambda}(t-K)}\mathbf{1}$
dies in $[r, r+1)$, $0 \le r \le K-1$	$\ddot{a}_R - \sum_{k=0}^{r} v^k$	$\underline{\tau}_R e^{\Lambda r}(\mathbf{1} - e^{\Lambda}\mathbf{1})$
dies in $[r, r+1)$, $K \le r \le t-1$	$\ddot{a}_R - \sum_{k=0}^{r} v^k$	$\underline{\tau}_R e^{\Lambda K} e^{\tilde{\Lambda}(r-K)}(\mathbf{1} - e^{\tilde{\Lambda}}\mathbf{1})$

Table 6.2: Individual present value $\psi^{\varepsilon}_{0,t}(K)$

where μ_K and θ^2_{K+1} are given in Theorem 6.1.1, $S_R(K)$ is defined in Eq. (2.4),

$$\ddot{a}^{\varepsilon}_{x:t} = \underline{\tau}_x (I - v^t e^{\tilde{\Lambda} t})(I - v e^{\tilde{\Lambda}})^{-1}\mathbf{1},$$

$$\Delta^{\varepsilon}_{t,K} = \sum_{i=1}^{t-K-1} \underline{\tau}_{R+K} e^{\tilde{\Lambda} i} v^i \mathbf{1}(\gamma_{i+K} - 2\ddot{a}_{R:K+1}) - S_R(K) v^K (\ddot{a}^{\varepsilon}_{R+K:t-K-1} - 1)^2.$$

Proof. We start with determining the expectation of $\psi^{\varepsilon}_{0,t}(K)$ from Table 6.2 as follows

$$\mu^{\varepsilon}_t = \left(\ddot{a}_R - \sum_{k=0}^{t-1} v^k\right) \underline{\tau}_R e^{\Lambda K} e^{\tilde{\Lambda}(t-K)}\mathbf{1} + \sum_{r=0}^{K-1}\left(\ddot{a}_R - \sum_{k=0}^{r} v^k\right)\underline{\tau}_R e^{\Lambda r}(I - e^{\Lambda})\mathbf{1}$$
$$+ \sum_{r=K}^{t-1}\left(\ddot{a}_R - \sum_{k=0}^{r} v^k\right)\underline{\tau}_R e^{\Lambda K} e^{\tilde{\Lambda}(r-K)}(I - e^{\tilde{\Lambda}})\mathbf{1}. \tag{6.25}$$

In Eq. (6.25) we take $\ddot{a}_R$ out, group the remainder of the first and the third terms to obtain

$$\begin{aligned}\mu^{\varepsilon}_t &= \ddot{a}_R - \underline{\tau}_R e^{\Lambda K}\left(\frac{1 - v^{K+1}}{1 - v} I + v^K \sum_{s=1}^{t-K-1} v^s e^{\tilde{\Lambda} s}\right)\mathbf{1} \\ &\quad - \underline{\tau}_R \sum_{r=0}^{K-1} \frac{1 - v^{r+1}}{1 - v} e^{\Lambda r}(I - e^{\Lambda})\mathbf{1}.\end{aligned}$$

We notice that $S_R(K)\underline{\tau}_{R+K} = \underline{\tau}_R e^{\Lambda K}$, since both parts of the equation are the probability to survive to age $R + K$, given the individual has survived to age R. Algebraically, the equation results from Eq. (2.3) and (2.4). Define $\ddot{a}^{\varepsilon}_{x:t}$ to be the annuity for t years for an individual aged x, calculated with the transition matrix $\tilde{\Lambda}$.

This gives

$$\begin{aligned}\mu_t^\varepsilon =&\ddot{a}_R - S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1)\\&-\frac{1-v^{K+1}}{1-v}\underline{\tau}_R e^{\Lambda K}\mathbf{1}-\sum_{r=0}^{K-1}\frac{1-v^{r+1}}{1-v}\underline{\tau}_R e^{\Lambda r}(I-e^{\Lambda})\mathbf{1}\\&=\ddot{a}_R - S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1)-\sum_{r=0}^{K}\underline{\tau}_R e^{\Lambda r}\mathbf{1}\left(\frac{1-v^{r+1}}{1-v}-\frac{1-v^{r}}{1-v}\right)\\&=\ddot{a}_R-(\ddot{a}_{R:K+1}+S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1))\\&=\mu_K - S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1).\end{aligned}\tag{6.26}$$

To determine the variance we start from the formula $\theta_t^{\varepsilon 2}=E[\psi^{\varepsilon}_{0,t}{}^2(K)]-(\mu_t^\varepsilon)^2$. To determine $E[\psi^{\varepsilon}_{0,t}{}^2(K)]$ we proceed in the same way as for Eq. (6.25), and we obtain

$$\begin{aligned}E[\psi_t^{\varepsilon 2}]=&\left(\ddot{a}_R-\sum_{k=0}^{t-1}v^k\right)^2\underline{\tau}_R e^{\Lambda K}e^{\tilde{\Lambda}(t-K)}\mathbf{1}+\sum_{r=0}^{K-1}\left(\ddot{a}_R-\sum_{k=0}^{r}v^k\right)^2\underline{\tau}_R e^{\Lambda r}(I-e^{\Lambda})\mathbf{1}\\&=\ddot{a}_R^2+\sum_{s=0}^{K}\underline{\tau}_R v^s e^{\Lambda s}\mathbf{1}(\gamma_s-2\ddot{a}_R)+v^K\sum_{s=1}^{t-K-1}\underline{\tau}_R v^s e^{\Lambda K}e^{\tilde{\Lambda}s}\mathbf{1}(\gamma_{s+K}-2\ddot{a}_R)\\&=\ddot{a}_R^2-2\ddot{a}_R(\ddot{a}_{R:K+1}+S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1))\\&\quad+\sum_{s=0}^{K}\underline{\tau}_R v^s e^{\Lambda s}\mathbf{1}\gamma_s+v^K\sum_{s=1}^{t-K-1}\underline{\tau}_R v^s e^{\Lambda K}e^{\tilde{\Lambda}s}\mathbf{1}\gamma_{s+K}.\end{aligned}$$

Therefore,

$$\begin{aligned}\theta_t^{\varepsilon 2}=&\ddot{a}_R^2-2\ddot{a}_R(\ddot{a}_{R:K+1}+S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1))\\&+\sum_{s=0}^{K}\underline{\tau}_R v^s e^{\Lambda s}\mathbf{1}\gamma_s+v^K\sum_{s=1}^{t-K-1}\underline{\tau}_R v^s e^{\Lambda K}e^{\tilde{\Lambda}s}\mathbf{1}\gamma_{s+K}\\&-(\ddot{a}_R-(\ddot{a}_{R:K+1}+S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1)))^2\\=&-(\ddot{a}_{R:K+1}+S_R(K)v^K(\ddot{a}^\varepsilon_{R+K:t-K-1}-1))^2\\&+\sum_{s=0}^{K}\underline{\tau}_R v^s e^{\Lambda s}\mathbf{1}\gamma_s+S_R(K)v^K\sum_{s=1}^{t-K-1}\underline{\tau}_{R+K}v^s e^{\tilde{\Lambda}s}\mathbf{1}\gamma_{s+K}\\=&-\ddot{a}_{R:K+1}^2+\sum_{s=0}^{K}\underline{\tau}_R v^s e^{\Lambda s}\mathbf{1}\gamma_s+S_R(K)v^K\sum_{i=1}^{t-K-1}\underline{\tau}_{R+K}e^{\tilde{\Lambda}i}v^i\mathbf{1}(\gamma_{i+K}-2\ddot{a}_{R:K+1})\\&-(S_R(K))^2v^{2K}(\ddot{a}^\varepsilon_{R+K:t-K-1}-1)^2=\theta^2_{K+1}+S_R(K)v^K\Delta^\varepsilon_{t,K}.\end{aligned}$$

■

Equation (6.26) shows that μ_t^ε can be interpreted as the difference between two annuities. The first $\ddot{a}_R$ represents accumulations accrued at retirement. The second is equal to the sum of two annuities of durations $K+1$ and $t-K-1$, respectively. The annuity of duration $K+1$ units of time is calculated with the original mortality matrix Λ, the annuity of duration $t-K-1$ units of time is calculated with the new mortality matrix $\tilde{\Lambda}$ and is adjusted to the present by discounting and multiplying by the probability to survive for K years.

Equation (6.24) shows that the variance of $\psi_{0,t}^\varepsilon(K)$ at time $t \geq K$ is the sum of two terms. The first is the accumulated variance at $t = K+1$ calculated in the original mortality assumptions; the second is $\Delta_{t,K}^\varepsilon$ adjusted to the present by discounting and multiplying by the probability to survive for K years.

For the model with Poisson arrivals, the expectation and the variance of ϕ_t^ε can be obtained from the following lemma.

Lemma 6.3.4 *The mean M_t^ε and the variance $\Theta_t^{\varepsilon 2}$ of ϕ_t^ε, $t \geq 1$ are*

$$
\begin{aligned}
M_t^\varepsilon &= \frac{1}{t}\sum_{k=0}^{t-1} \mu_{t-k}^\varepsilon v^k \\
\Theta_t^{\varepsilon 2} &= \frac{1}{t}\sum_{k=0}^{t-1} (\theta_{t-k}^\varepsilon)^2 v^{2k} + \frac{1}{t}\sum_{k=0}^{t-1} (\mu_{t-k}^\varepsilon)^2 v^{2k} - \left(\frac{1}{t}\sum_{k=0}^{t-1} \mu_{t-k}^\varepsilon v^k\right)^2,
\end{aligned}
$$

where μ_{t-k}^ε and $(\theta_{t-k}^\varepsilon)^2$ are given by Eq. (6.23) *and Eq.* (6.24)*, respectively.*

We give an illustrative example in Fig. 6.7, where we depict $E[V_t^\varepsilon]$ for $\varepsilon = 0$, 0.1 and 0.2 and also $E[V_t^*]$ with retirement at age 62. We observe that $E[V_t^*]$ is almost identical to $E[V_t^{0.2}]$, which indicates that the overall reduction of mortality rates by 20% has approximately the same financial impact as retirement at age 62 instead of 65. Therefore, we remark that the constructed model also allows us to compare the impacts of mortality reduction and of other important events – *early retirement* in this case.

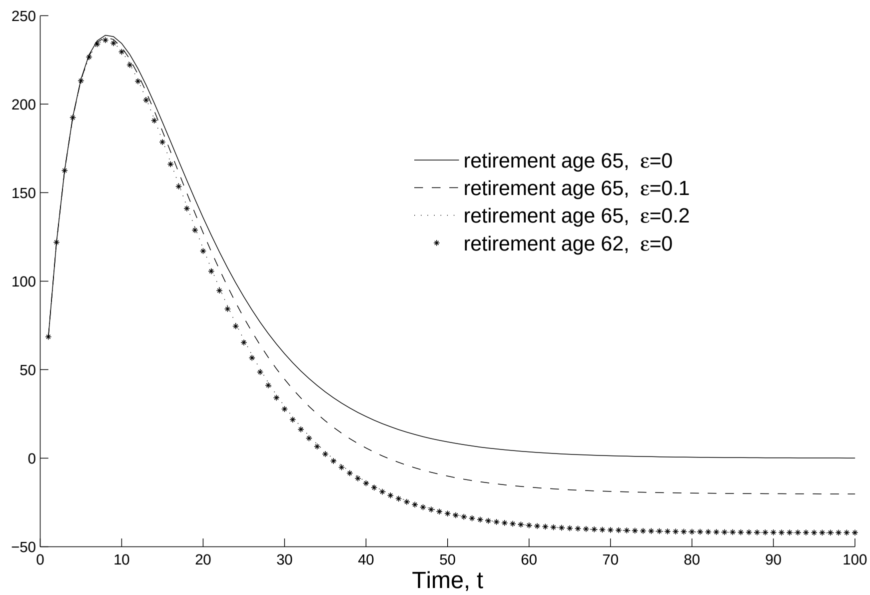

Figure 6.7: Impact of external factors: $E[V_t^{\varepsilon}]$ and $E[V_t^*]$
Parameters: SW1911M, $N = 10$, $v = 0.9091$

Part III

Health Care

Summary of Part III

The connection, provided by the phase-type lifetime assumption, between the age and health opens new horizons in health insurance. This part is devoted to the estimation of health care costs for an individual of a given age.

Modeling health care costs is a problem of great interest in health insurance and economics (Zhao and Zhou [67]). The estimations of the costs play a key role in pricing, reserving and risk assessment in health insurance, as well as performing cost-effectiveness and cost-utility analyses in health economics. The modeling of health care costs is not a simple problem as the accumulated cost by an individual is correlated with his/her survival time.

Often it is assumed that health care costs depend on an individual health state, which is modeled by a Markov chain. In Castelli et al. [17], Gardiner et al. [25] and Zhao and Zhou [67] the Markov chain has a fixed number of states, which is a subjectively chosen parameter, and does not depend on the age of an individual. The health care cost related quantities that are studied in [17], [25] and [67] are slightly different as we showed in Section 1.4, but all of them are computed as expected values. In Gardiner et al. [25], the authors work in continuous time and determine the expected *net present value* (abbreviated as "NPV" in the sequel) of health care costs over a fixed time horizon.

In Chapter 7 we are focused not only on the expectation of NPV, but also on the distribution. We develop three models for NPV with our underlying assumption being that the lifetime and health of an individual are described by the PH-type aging model with the generator Q (see Eq. (2.1)) that has a two-diagonal or any structure.

The first two models are discrete time models designed to determine the distribution of NPV over a long-term horizon such as the lifetime of an individual, which is a phase-type random variable. In the first model (called "model with iid costs") we assume that the annual health care costs are independent and identically distributed random variables, and thus they do not depend on the aging of an individual. Indeed, in reality not all types of health care costs depend on the aging process. There is almost no such dependence for dental care costs as well as for costs related to a health damage caused by accidents. In insurance, this approach might be useful to estimate *personal accident* type of policies. In the United States, for instance, there is a *personal injury protection* policy in auto insurance, which covers medical and funeral expenses associated to a car accident and which is obligatory in some of the states. In the second model (called "Markov reward model for costs"), similarly to [17], [25] and [67], we assume that the annual health care costs depend on the aging process and that the value of the cost is a constant for each given state. In the two models, NPV is a compound random variable, where the number of terms in the sum has a phase-type distribution and where a time-dependent coefficient is applied to each term. In the absence of this coefficient and if the health costs are iid, to find the distribution of NPV one may apply the algorithm suggested by Eisele [22], which is a natural extension of the widely used procedure introduced by Panjer [49]. In our work, we develop a method that enables us to write similar algorithms for the distribution of NPV in both models. We compare the two models by providing a simple parametrization procedure, and we perform a sensitivity analysis to study the effect of different phenomena. For example, tests with respect to mortality rates allow us to estimate the impact of an increased lifetime spent in bad health states for which medical treatments are the most expensive.

The third model is a continuous-time model designed to compute the distribution of NPV over a fixed short-term horizon. The motivation is that regulatory requirements often impose restrictions on a short-term basis, implying that the risks have to be calculated with a good precision at any time. The model represents a continuous time version of the Markov reward model for costs and employs a *fluid queue* with time-dependent rates to describe the behavior of NPV in time.

In Chapter 8 we extend the discrete- and continuous-time models with health state dependent costs by allowing the cost to be random for a given health state. In discrete time, the model, called "Randomized Markov reward model", is similar to the Markov reward model for costs, the difference being that the cost for a given state is no longer a constant, but is a random variable with a distribution defined by the state. In continuous time, we suggest two models to introduce the uncertainty in the cost for a given state. In both models NPV is represented by a fluid queue with time-dependent stochastic rates. In the first model (called "fluid model with geometric cost rates") the cost rate for a given state is defined as a *geometric Brownian motion* with parameters that depend on the state. The geometric Brownian motion is widely used in mathematical finance to model non-negative random variables such as prices or interest rates (see, for example, Lin [39]). In the second model (called "fluid model with Brownian increments") the cost rates are given by the increments of a *Brownian motion* with state-dependent parameters, a process widely used in ruin theory and in mathematical finance, see Section 1.4.

The principal difference between the two models is illustrated in Fig. 8.2 and Fig. 8.3: in the first model the cost rate has a positive drift for a given state; in the second model the cost rate fluctuates around its average constant value. Another important difference between the two models is that the fluid model with Brownian increments allows the cost rates to be negative. The assumption that the costs can be of any sign might be useful in several cases. For example, if at time u an individual receives from an insurance institution a health treatment coverage, which has been overestimated. In this case, at time $u_1 > u$ the part of the coverage that exceeded real treatment costs may be reimbursed to the insurance institution. If at time u_1 the individual did not need any treatment, the total cost rate at time u_1 is obviously negative. Another example is the situation, when the cost rates are re-defined as a total cash flow and contain also regular contributions of the individual to his/her health insurance plan. In other words, the model can be used to estimate the NPV of future profits and losses. If one needs only non-negative costs, we indicate how the fluid model with Brownian increments can be adapted. There are several advantages of the model over the fluid model with geometric cost rates that we underline at the end of the chapter.

As a main contribution of Part III we indicate the development of various mathematical models to determine recursive equations for the distribution of NPV. Here, the lifetime and health of the individual are described by the PH-aging model with the generator Q (see Eq. (2.1)) of a two-diagonal or any structure.

In discrete time, we have elaborated a model with iid costs and models with health dependent costs with constant and random costs for a given state. We have obtained recursive equations for the distribution of NPV over a phase-type time horizon by extending the procedure introduced in Panjer [49]. For each of the models, we have also derived closed form expressions for the expectation of NPV. We have suggested a parametrization procedure for the model with iid costs and for the model with health dependent constant costs. As a result of this procedure we have obtained a specific parametrization under which the model with iid costs gives approximately same expected value of NPV as the model with health dependent costs. Furthermore, we have elaborated a technique to estimate the financial impact of longer survival in bad health states, which we have applied to the model with health dependent constant costs.

In continuous time, we have elaborated three fluid models with health dependent costs rates. In the first fluid model, the cost rate is a constant for each given health state, and in the two other models the cost rate is a state-dependent random variable. In the last two models, the cost rate for a given state is defined by a stochastic process, a geometric Brownian motion or an instantaneous increment of a Brownian motion, with parameters that depend on the state. We have obtained recursive equations for the distribution of NPV over a fixed time horizon by using a renewal argument. To apply the renewal argument in the models with stochastic cost rates, we have additionally derived the distribution of the increments of NPV for a given state. For each of the three fluid models we have developed two sets of recursive equations: a simpler one when Q has a two-diagonal structure, and a more complex one when Q is a generator of any structure.

Needless to remark that as far as we know fluid queues with time-dependent deterministic or stochastic rates have not yet been well studied in the literature. Therefore, the equations that we have obtained for the distribution of NPV ("level" of a fluid queue) in our continuous time models may be considered as a novel contribution to the theory of fluid queues.

Another remark is that the developed models and equations are general enough and may be applied to any other actuarial quantities where total discounted payments are correlated with the lifetime of an individual.

Chapter 7

Net present value of health care costs

In this chapter we deal with the *net present value of costs* in health insurance. In Section 7.1 we start with the introduction of the problem and some algorithmic aspects that help us to determine the distribution of the net present value.

In order to evaluate the distribution over a long-term horizon, we define two discrete time models that are based on different assumptions for health care costs. In Section 7.2 we assume that annual individual health care costs are independent and identically distributed random variables for each year of life. Thus, they do not depend on the aging of individuals. In Section 7.3 we present our second model, where we assume that health care costs depend on the health state of the individual and that they are defined by a *Markov reward process* introduced in Section 5.4. We discuss the aspects of the parametrization and sensitivity analysis of the models in Sections 7.4 and 7.5, respectively.

In Section 7.6 we develop a continuous time model, where the net present value of health care costs is described by a *fluid queue*.

7.1 Main objectives

Our main objective is to develop a phase-type method to obtain the distribution of the net present value of a health care contract. In continuous time the *net present* *value* is given by

$$S_t = \int_0^t v^u X_u du \quad \text{or} \quad S = \int_0^L v^u X_u du, \tag{7.1}$$

where v is a constant coefficient, X_u is a health care cost rate at time u, t is a fixed time horizon. Here and below, we assume that the individual is of age x at time zero, and we denote his/her remaining lifetime by L, omitting the index x in the notation.

The remaining lifetime L of the individual is a continuous time random variable, resulting from the *PH-aging model*. Thus, $L \sim PH(\underline{\tau}, \Lambda)$, where $\underline{\tau}$ is the health state distribution at age x, given by Eq. (2.3), and Λ is the aging transition rate matrix, defined by Eq. (2.2). The expected value of L is given by Eq. (1.28), where $\underline{\alpha}$ is replaced by $\underline{\tau}$.

For a life long duration we adopt a discrete time approach and define the net present value of a health care contract as

$$S = \sum_{t=1}^{[L]} v^{t-1} X_t, \tag{7.2}$$

where $[L]$ is the integer number of remaining life years, and X_t is the health care cost in year t. The coefficient v is allowed to take any positive value and v including greater than one, so as to include inflation, interest force, the increase of health care prices, etc.

Define by ψ_t the health state of the individual at time t. As seen in Section 2.1, the probability to survive for t years and to be in state i at time t is $\left(\underline{\tau} e^{\Lambda t}\right)_i$, $i = 1, .., n$. We aim to determine the distribution of S, which we denote by $G_S(k)$. The problem is equivalent to the problem of finding the conditional distribution given the health state at time 0, because $G_S(k) = \underline{\tau}\, \underline{H}^{\mathrm{T}}(k)$, where

$$\underline{H}(k): \ H_i(k) = \mathrm{P}\left[\, S \le k \mid \psi_0 = i \,\right], \quad i = 1, ..., n. \tag{7.3}$$

In order to obtain a recursive equation for $\underline{H}(k)$ we notice that

$$\text{if} \quad [L] = 1 \quad \text{then} \quad S = X_1, \tag{7.4}$$

$$\text{if} \quad [L] \ge 2 \quad \text{then} \quad S = X_1 + v\tilde{S}, \tag{7.5}$$

where $\tilde{S}$ is a random variable with the same transition matrix as S, but with a different initial health state vector. Denote by $\underline{y}$ the conditional probability to die in any given year, given the health state at time 0. It is equal to one minus one year survival probability, so that

$$\underline{y} = \mathbf{1} - e^{\Lambda}\mathbf{1}. \tag{7.6}$$

We determine the exact form of the recursive equations for $\underline{H}(k)$ in Sections 7.2 and 7.3, where we consider two different models for health care costs, X_t.

In the short-term, when t is not big, we aim to compute the net present value precisely at any time, therefore it is convenient to adopt a continuous time approach. In this approach, we determine the distribution of S_t defined by Eq. (7.1). In order to obtain the distribution, in Section 7.6 we apply a renewal argument similar to the one given by Eq. (7.4) and Eq. (7.5).

7.2 Model with iid costs

For some types of health care costs in discrete time it is reasonable to assume that annual health care costs X_t are random, independent and have the same distribution. This assumption leads to the following theorem. In continuous time, we defer the discussion to Section 8.4.

Theorem 7.2.1 ***(iid costs)*** *Suppose that X_t are discrete iid random variables that take M non-negative possible values, $X_t \in \{c_1, ..., c_M\}$, with $F(k) = P[X_t \leq k]$ and $f(k) = P[X_t = k]$. The conditional distribution $\underline{H}(k)$ of the net present value S of a health care contract defined by Eq. (7.2) is such that*

$$H_i(k) = y_i F(k) + \left(e^{\Lambda}\right)_{(i,\cdot)} \sum_{\theta=1}^{M} f(c_\theta) \underline{H}^T \left(\frac{k - c_\theta}{v} \right), \quad k \geq 0, \tag{7.7}$$

where $\underline{y}$ is given by Eq. (7.6), $i = 1, ..., n$. In particular,

$$H_i(0) = f(0) \underline{\alpha}^{(i)} e^{\Lambda} (I - f(0) e^{\Lambda})^{-1} \underline{y}^T, \tag{7.8}$$

where $\underline{\alpha}^{(i)}$ is defined in Eq. (2.11).

Proof. The proof is based on the representation of S given by Eq. (7.4) and Eq. (7.5):

$$\begin{aligned}H_i(k) =&\mathrm{P}\,[\,[L]=1\,|\,\psi_0=i\,]\mathrm{P}[\,X_1\leq k\,]\\ &+\sum_{j\in\{A\}}\mathrm{P}[\,\psi_1=j\,|\,\psi_0=i\,]\mathrm{P}[\,X_1+v\tilde{S}\leq k\,|\,\psi_1=j\,]\\ =&y_iF(k)+\sum_{j\in\{A\}}\mathrm{P}[\,\psi_1=j\,|\,\psi_0=i\,]\sum_{\theta=1}^{M}f(c_\theta)H_j\left(\frac{k-c_\theta}{v}\right)\\ =&y_iF(k)+\sum_{j\in\{A\}}(e^{\Lambda})_{(i,j)}\sum_{\theta=1}^{M}f(c_\theta)H_j\left(\frac{k-c_\theta}{v}\right)\\ =&y_iF(k)+\left(e^{\Lambda}\right)_{(i,\cdot)}\sum_{\theta=1}^{M}f(c_\theta)\underline{H}^{\mathrm{T}}\left(\frac{k-c_\theta}{v}\right).\end{aligned}$$

If $k=0$, this implies that all costs X_t, $t=1,...,[L]$ are equal to zero and, therefore,

$$\begin{aligned}H_i(0)=&\sum_{s=1}^{\infty}\mathrm{P}[\,[L]=s\,](f(0))^s=\sum_{s=1}^{\infty}\underline{\alpha}^{(i)}e^{\Lambda s}\underline{y}^{\mathrm{T}}(f(0))^s\\ =&\underline{\alpha}^{(i)}f(0)e^{\Lambda}\sum_{s=0}^{\infty}e^{\Lambda s}\underline{y}^{\mathrm{T}}(f(0))^s\\ =&f(0)\underline{\alpha}^{(i)}e^{\Lambda}(I-f(0)e^{\Lambda})^{-1}\underline{y}^{\mathrm{T}}.\end{aligned}$$

■

Lemma 7.2.2 *The expectation of S in the model with iid costs X_t is given by*

$$E[S]=\underline{\tau}(I-ve^{\Lambda})^{-1}\mathbf{1}E[X_t]. \tag{7.9}$$

Proof. The expectation of S is computed as follows

$$E[S]=E\left[\sum_{t=1}^{[L]}v^{t-1}X_t\right]=E[X_t]E\left[\frac{1-v^{[L]}}{1-v}\right]=\frac{1}{1-v}(1-g(v))E[X_t], \tag{7.10}$$

where $g(v)$ is the generating function of a discrete phase-type random variable defined in Eq. (1.30) as

$$g(v)=v\underline{\tau}(I-ve^{\Lambda})^{-1}(\mathbf{1}-e^{\Lambda}\mathbf{1}).$$

One can represent 1 as $\underline{\tau}(I - ve^{\Lambda})^{-1}(I - ve^{\Lambda})\mathbf{1}$. Therefore,

$$\begin{aligned} 1 - g(v) &= \underline{\tau}(I - ve^{\Lambda})^{-1}(I - ve^{\Lambda})\mathbf{1} - v\underline{\tau}(I - ve^{\Lambda})^{-1}(\mathbf{1} - e^{\Lambda}\mathbf{1}) \\ &= \underline{\tau}(I - ve^{\Lambda})^{-1}(I - ve^{\Lambda} - vI + ve^{\Lambda})\mathbf{1} \\ &= (1-v)\underline{\tau}(I - ve^{\Lambda})^{-1}\mathbf{1}. \end{aligned}$$

This result together with Eq. (7.10) gives the statement of the lemma. ■

7.3 Markov reward model for costs

We capture the dependence between the health care costs and the aging of an individual by introducing a Markov reward process. In this framework it is defined as the triplet $(A \cup \{D\}, \underline{W}, P)$, where

i. $A \cup \{D\}$ is the set of possible states: A is the set of n health states from the aging model, $\{D\}$ is one absorbing state for death;

ii. $\underline{W}$ is the vector of size $n+1$ of health care costs for each state so that

$$W_i \geq 0, \text{ if } i \in A, \quad \text{and} \quad W_i = 0, \text{ if } i \in \{D\}; \tag{7.11}$$

iii. P is the state transition matrix, which is constructed from the generator of the aging model as follows

$$P = \begin{bmatrix} e^{\Lambda} & \underline{y}^{\mathrm{T}} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}, \tag{7.12}$$

where $\underline{y}$ is given by Eq. (7.6). Denote the corresponding health state process by ϕ_t, $t = 0, 1, 2, ...$

The application of the Markov reward process results in a simplier equation for $\underline{H}(k)$, as shown by the following theorem.

Theorem 7.3.1 *Suppose that X_t follows the Markov Reward process defined by the triplet $(A \cup \{D\}, \underline{W}, P)$. The conditional distribution $\underline{H}(k)$ of the net present value S of a health care contract defined by Eq.* (7.2) *is such that*

$$H_i(k) = y_i \mathbb{1}_{\{W_i \leq k\}} + (e^{\Lambda})_{(i,\cdot)} \underline{H}^T\left(\frac{k - W_i}{v}\right), \quad k \geq 0, \tag{7.13}$$

where $\underline{y}$ is given by Eq. (7.6), *$i = 1, .., n$.. In particular,*

$$\text{if } W_i \neq 0 \quad \text{then} \quad H_i(0) = 0, \tag{7.14}$$

$$\text{if } W_i = 0 \quad \text{then} \quad H_i(0) = y_i + (e^{\Lambda})_{(i,\cdot)} \underline{H}^T(0). \tag{7.15}$$

Proof. The proof is analogous to the proof of Theorem 7.2.1. To show the statement we use the fact that $\mathrm{P}[\, X_1 \leq k \,|\, \phi_0 = i\,] = \mathbb{1}_{\{W_i \leq k\}}$ and $\mathrm{P}[\, X_1 = k \,|\, \phi_0 = i\,] = \mathbb{1}_{\{W_i = k\}}$, because X_1 is deterministic given the health state i at the beginning of the year. ■

When $k = 0$ Eq. (7.14) and Eq. (7.15) show that Eq. (7.13) is no longer a recursive equation, but it defines a linear system. The following lemma gives a closed form expression for $E[S]$, which is similar to the one given by Lemma 7.2.2 for the iid model. Specifically, in Eq. (7.9) the vector $\mathbf{1}E[X_t]$ is replaced by $\underline{W}^{\mathrm{T}}$.

Lemma 7.3.2 *The expectation of S for the Markov reward model $(A \cup \{D\}, \underline{W}, P)$ is given by*

$$E[S] = [\underline{\tau}\ 0](I - vP)^{-1}\underline{W}^{T}. \tag{7.16}$$

Proof. Denote by $g(\zeta)$ the probability generating function of S: $g(\zeta) = E[\zeta^S]$. We have

$$g(\zeta) = [\underline{\tau}\ 0]\underline{f}^{\mathrm{T}}(\zeta), \tag{7.17}$$

where $\underline{f}^{\mathrm{T}}(\zeta)$ is a column vector of size $n+1$ such that $f_i(\zeta) = E[\,\zeta^S \,|\, \phi_0 = i\,]$, $i \in A \cup \{D\}$. If $[L] = 1$, then by Eq. (7.4) $f_i(\zeta) = \zeta^{W_i}$. If $[L] \geq 2$, by conditioning on the first transition, we obtain

$$\begin{aligned} f_i(\zeta) &= \sum_{j \in A \cup \{D\}} \mathrm{P}[\,\phi_1 = j \,|\, \phi_0 = i\,]\, E[\,\zeta^S \,|\, \phi_0 = i, \phi_1 = j\,] \\ &= \sum_{j \in A \cup \{D\}} \mathrm{P}[\,\phi_1 = j \,|\, \phi_0 = i\,]\, E[\,\zeta^{W_i + v\tilde{S}} \,|\, \phi_0 = i, \phi_1 = j\,] \\ &= \zeta^{W_i} \sum_{j \in A \cup \{D\}} \mathrm{P}[\,\phi_1 = j \,|\, \phi_0 = i\,]\, E[\,(\zeta^v)^{\tilde{S}} \,|\, \phi_1 = j\,] \\ &= \zeta^{W_i} \sum_{j \in A \cup \{D\}} P_{(i,j)}\, E[\,(\zeta^v)^S \,|\, \phi_0 = j\,] \\ &= \zeta^{W_i} \sum_{j \in A \cup \{D\}} P_{(i,j)} f_j(\zeta^v). \end{aligned} \tag{7.18}$$

In the matrix form, Eq. (7.18) is written as

$$\underline{f}^{\mathrm{T}}(\zeta) = \Upsilon P \underline{f}^{\mathrm{T}}(\zeta^v), \tag{7.19}$$

where Υ is a diagonal matrix of size $n+1$, given by

$$\Upsilon = \begin{bmatrix} \zeta^{W_1} & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \zeta^{W_n} & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}. \tag{7.20}$$

We employ the properties of generating functions and Eq. (7.17) to obtain

$$E[S] = \left.\frac{\partial g(\zeta)}{\partial \zeta}\right|_{\zeta=1} = [\underline{\tau}\ 0]\left.\frac{\partial \underline{f}^{\mathrm{T}}(\zeta)}{\partial \zeta}\right|_{\zeta=1}. \tag{7.21}$$

Denote

$$\widetilde{W} = \begin{bmatrix} W_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & W_n & 0 \\ 0 & \dots & 0 & 0 \end{bmatrix}. \tag{7.22}$$

By differentiating Eq. (7.19), we find

$$\begin{aligned} \frac{\partial \underline{f}^{\mathrm{T}}(\zeta)}{\partial \zeta} &= \frac{1}{\zeta}\widetilde{W}\Upsilon P \underline{f}^{\mathrm{T}}(\zeta^v) + v\zeta^{v-1}\Upsilon P \left.\frac{\partial \underline{f}^{\mathrm{T}}(z)}{\partial z}\right|_{z=\zeta^v} \\ &= \frac{1}{\zeta}\widetilde{W}\underline{f}(\zeta) + v\zeta^{v-1}\Upsilon P \left.\frac{\partial \underline{f}(z)}{\partial z}\right|_{z=\zeta^v}. \end{aligned} \tag{7.23}$$

We evaluate Eq. (7.23) at $\zeta = 1$ and use $\underline{f}^{\mathrm{T}}(1) = \mathbf{1}$ to obtain

$$\left.\frac{\partial \underline{f}^{\mathrm{T}}(\zeta)}{\partial \zeta}\right|_{\zeta=1} = \widetilde{W}\mathbf{1} + vP\left.\frac{\partial \underline{f}^{\mathrm{T}}(z)}{\partial z}\right|_{z=1}, \tag{7.24}$$

and, after rearranging the terms,

$$\left.\frac{\partial \underline{f}^{\mathrm{T}}(\zeta)}{\partial \zeta}\right|_{\zeta=1} = (I - vP)^{-1}\underline{W}^{\mathrm{T}}. \tag{7.25}$$

By combining Eq. (7.21) and Eq. (7.25) we obtain the statement of the lemma. ■

7.4 Parametrization aspects

In order to indicate how the model with iid costs and the Markov reward model can be used in practice, we give here one example of a parametrization procedure.

What is the available data? This is the first natural question that appears when one intends to parametrize a model. We may assume that health care costs are available for each age, from 0 to x_{max} (100 in our examples), possibly as frequency

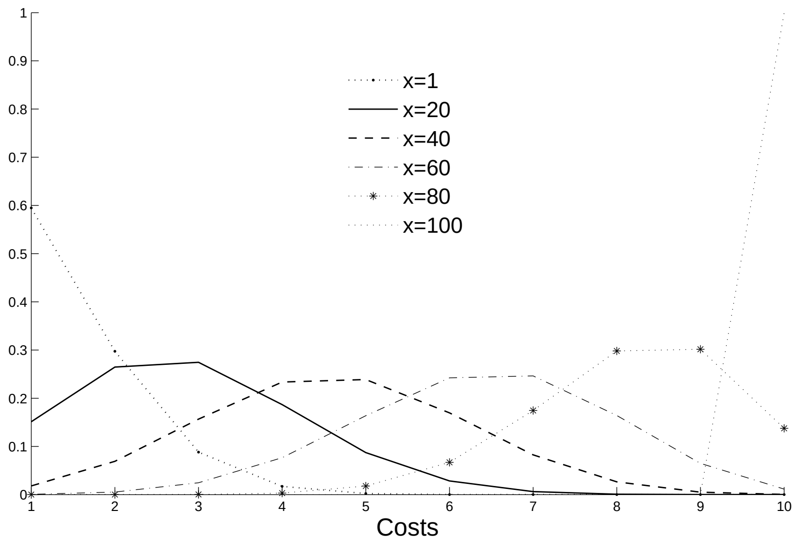

Figure 7.1: Distributions of Binomial $\hat{Y}_x$, $x =$ 1-100, constrained in $[1, 10]$

distributions and more likely as averages only. We denote by $\hat{Y}_x$ the health care cost for age x, observed from the data and we denote by M an integer such that $\hat{Y}_x \leq M$ with probability one for all x.

On the illustrative example to follow, we take $M = 10$ and assume that $\hat{Y}_x$ has $E[\hat{Y}_0] = 1$, $E[\hat{Y}_{100}] = 10$ and $E[\hat{Y}_x]$ is linear in between. To give a better feeling for the distributions to follow, we plot in Fig. 7.1 an example, when $\hat{Y}_x$ has a Binomial distribution with the probability mass constrained to be between 1 and 10. In our numerical examples we consider an individual aged 40 at time zero.

Markov reward model for costs. In order to parametrize the model, where health care costs follow a Markov reward process, we implement the procedure described in Section 2.3, and so we determine a value of the cost W_i for every health state $i \in A$. To obtain the values we assume that the costs are non-decreasing with respect to health, and we require the equality of their first moments for all ages. Namely, we solve numerically the linear optimization problem with constraints

$$\min_{\underline{W}_A} \|\mathcal{T}\underline{W}_A^{\mathrm{T}} - \mathcal{J}\|_2^2, \quad \underline{W}_A : \ 0 < W_i \leq W_{i+1} \leq M, \ \ i = 1, ..., n-1 \tag{7.26}$$

Here, $\mathcal{T} = [\underline{\tau}_x]_{x=1,..,100}$ and $\mathcal{J} = [E[\hat{Y}_x]]_{x=1,..,100}$. The result of the minimization is presented in Fig. 7.2. We recognize that the health costs for the states have a piece-wise constant structure similar to the one shown in Fig. 2.8. As one can see

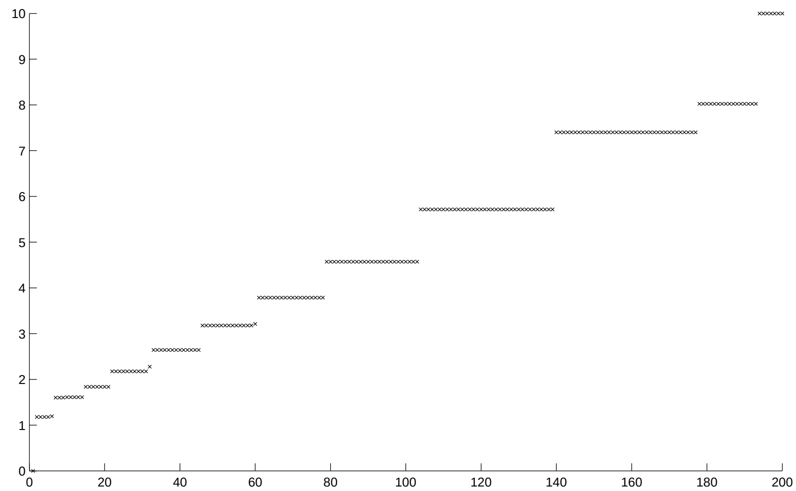

Figure 7.2: Health care costs W_i for states $i \in \{A\}$
Markov reward model, $M = 10$, SW1911M

in Fig. 7.3, the fit of average costs for ages is quite good, especially for young ages. The total sum of the squared differences between the average costs is 0.6242 with maximum cost value being 10.

Model with iid costs. For this model, we need a common distribution for the costs X_t of successive years. We denote by X a random variable with this common distribution. An individual aged x in the beginning of a health care contract can potentially live until age x_{max}, which signifies that X might be considered as a combination of health care costs for future years of life:

$$X = \sum_{j=x}^{x_{max}} \beta_{j-x} \hat{Y}_j, \tag{7.27}$$

where β_j are weights reflecting "the importance of each year". The weights β_j can be chosen in many different ways. We have considered the three following approaches:

i. "All years have the same value". In this case, the weights are uniform,

$$\beta_j = \frac{1}{x_{max} - x}, \tag{7.28}$$

and we denote by $X^{(1)}$ and $S^{(1)}$ the values of X and S, respectively;

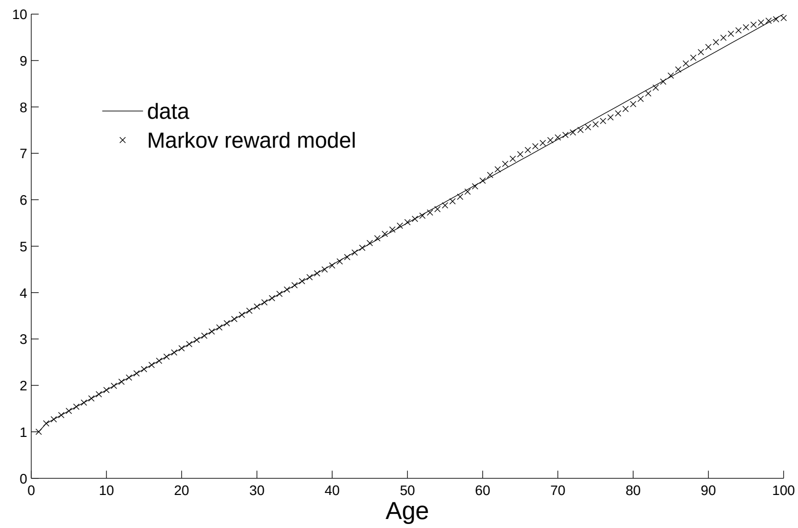

Figure 7.3: Expected health care costs for ages 1-100
$M = 10$, SW1911M

ii. "Later years are less representative because fewer individuals survive". Here, the weights are

$$\beta_j = \frac{S_x(j)}{\sum\limits_{j\geq 0} S_x(j)}, \tag{7.29}$$

where $S_x(j)$ is the j years survival probability defined by Eq. (2.4). We denote by $X^{(2)}$ and $S^{(2)}$ the values of X and S, respectively;

iii. "The importance of future years depends on the survival probabilities and on economical factors". Here, the weights are given by

$$\beta_j = \frac{v^j S_x(j)}{\sum\limits_{j\geq 0} v^j S_x(j)}, \tag{7.30}$$

where v is the coefficient in Eq. (7.2). If $v < 1$, future years have less importance; if $v > 1$, they get more weight. We denote by $X^{(3)}$ and $S^{(3)}$ the values of X and S, respectively.

In Fig. 7.4 we illustrate the three different choices of the distribution for X. The dashed lines in the figure represent our toy data: the distributions of health care

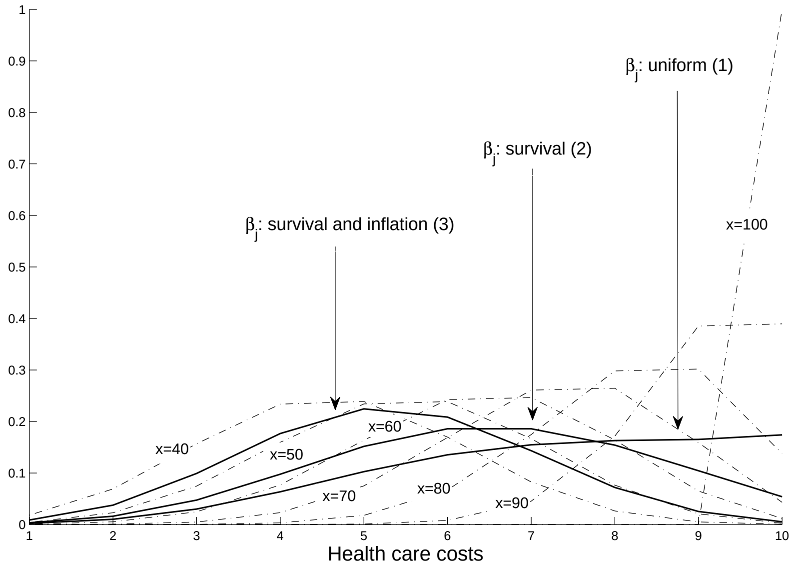

Figure 7.4: Distribution of X_i in the iid model for costs
$v = 0.9$, $x_{max} = 100$, $\hat{Y}_x$ from Fig. 7.1, SW1911M

costs $\hat{Y}_x$ for different values of x, from 40 to 100 at intervals of 10 years. The three solid lines stand for the distribution of X for the three choices of coefficients β_j. One observes from the figure that the choice of β_j has a significant impact on the distribution of X.

Remark 7.4.1 *If the weights β_j, $j = 0, .., x_{max} - x$ are chosen according to* (7.30)*, then the Markov reward model and the model with iid costs give approximately the same expectation of S. The higher the value of x_{max}, the lower the difference.*

Indeed, in the Markov reward model for costs, the expectation of S is

$$E[S] = \sum_{i \geq 0} v^i \underline{\tau}_x e^{\Lambda i} \underline{W}_A^{\mathrm{T}} = \sum_{i \geq 0} v^i \underline{\tau}_x e^{\Lambda i} \mathbf{1} \underline{\tau}_{x+i} \underline{W}_A^{\mathrm{T}} \tag{7.31}$$

$$= \sum_{i \geq 0} v^i S_x(i) (\mathcal{T} \underline{W}_A^{\mathrm{T}})_{x+i}, \tag{7.32}$$

where $(\mathcal{T}\underline{W}_A^{\mathrm{T}})_j$ is the average health cost at age j from Eq. (7.26).

In the model with iid costs and weights β_j chosen as in Eq. (7.30) the expectation of S is

$$E[S^{(3)}] = E[X^{(3)}] \sum_{i \geq 0} v^i S_x(i) \tag{7.33}$$

$$= \left(\sum_{j=0}^{x_{max}-x} \beta_j E[\hat{Y}_{j+x}] \right) \left(\sum_{i \geq 0} v^i S_x(i) \right)$$

$$= \sum_{j=0}^{x_{max}-x} v^j S_x(j) E[\hat{Y}_{x+j}]. \tag{7.34}$$

In Eq. (7.32) and Eq. (7.34) the terms $(\mathcal{T}\underline{W}_A^{\mathrm{T}})_{x+j}$ and $E[\hat{Y}_{x+j}]$ are only approximately the same as a result from the minimization procedure (7.26). A second slight difference between the equations comes from the upper summation limits: whereas in Eq. (7.32) the summation goes up to infinity, in Eq. (7.34) it goes only up to $(x_{max} - x)$. Recall that x_{max} is the maximum observed age. This indicates that for an individual aged x the probability to survive for more than $(x_{max} - x)$ years should be very small, and this small probability determines the difference due to truncation. Furthermore, the higher the x_{max}, provided in the observations, the lower is the difference. Thus, we have the statement of the lemma.

Remark 7.4.1 provides a useful property of the choice number 3, which we consider as the best of the three. Another interesting remark concerning the choice of coefficients β_j is given below.

Remark 7.4.2 *The ratio $E[S^{(1)}]/E[S^{(2)}]$ of the expected values of S with the weights given by* (7.28) *and by* (7.29) *is invariant with respect to v.*

Indeed, Eq. (7.33) indicates that $E[S]$ in both cases is the product of $E[X]$ and $\sum_{i \geq 0} v^i S_x(i)$. The first factor does not depend on v, because the weights β_j, given by (7.28) and (7.29), do not depend on v. The second factor does not depend on the weights. Thus, $E[S^{(1)}]/E[S^{(2)}] = E[X^{(1)}]/E[X^{(2)}]$ and it is independent of v.

We present in Fig. 7.5 and in Fig. 7.6, respectively, the distributions of S, calculated with $v = 0.7$ and with $v = 1.02$. Let us denote by S^* the net present value S in the Markov reward model. By looking at the figures one notices that the distribution of S^* and the distribution of $S^{(3)}$ are close to each other. The computed expectations of S are: $E[S^*] = 15.91$ and $E[S^{(3)}] = 15.93$ for $v = 0.7$, and $E[S^*] = 413.16$ and $E[S^{(3)}] = 407.5$ for $v = 1.02$, which confirms Remark 7.4.1. We also notice that in both figures $S^{(1)}$ is stochastically greater than $S^{(2)}$. This is due to the fact that the

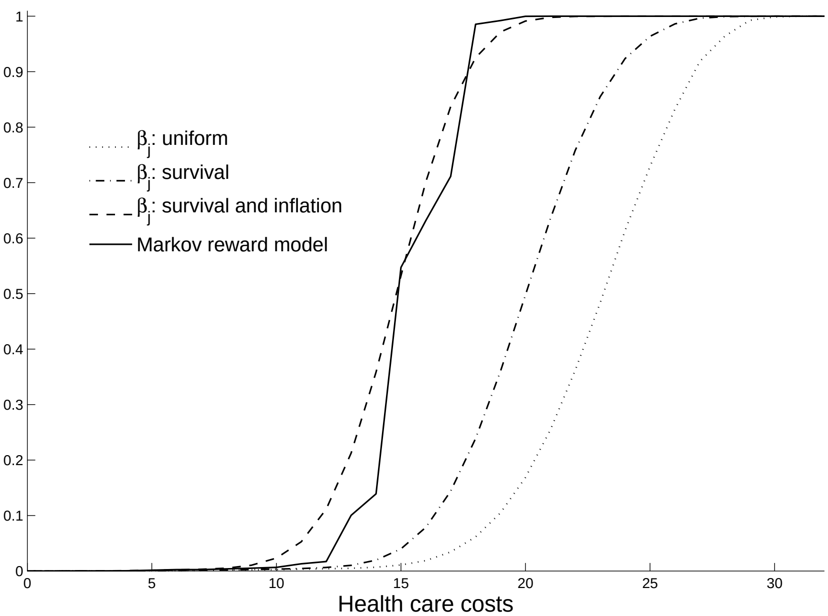

Figure 7.5: Distribution of S, $v = 0.7$, $x_{max} = 100$
W_i, $i \in \{A\}$ from Fig. 7.2, $\hat{Y}_x$ from Fig. 7.1, SW1911M

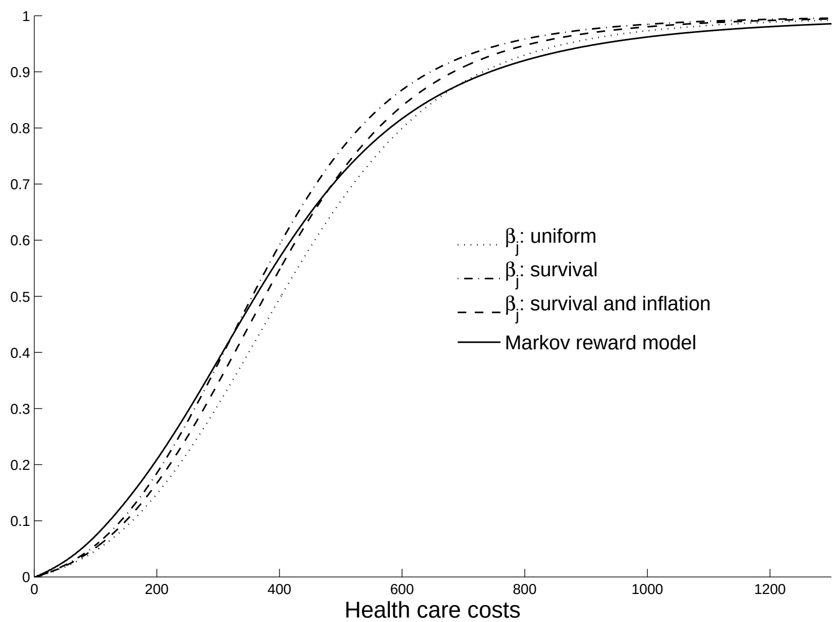

Figure 7.6: Distribution of S, $v = 1.02$, $x_{max} = 100$
W_i, $i \in \{A\}$ from Fig. 7.2, $\hat{Y}_x$ from Fig. 7.1, SW1911M

expected costs $E[\hat{Y}_x]$ increase with x and that later ages receive more weight in $X^{(1)}$ than in $X^{(2)}$.

By examining Fig. 7.5 for $v = 0.7$, we conclude that the choice of the mixing factors β_j significantly affects the distribution of S. In Fig. 7.6 for $v = 1.02$ the distributions of $S^{(1)}, S^{(2)}$ and $S^{(3)}$ seem not to differ very much; this is mostly due to the fact that for v greater than one the distribution of X becomes similar for all three choices of β_j, and because v becomes the dominant effect (the x-axis scale is different for Fig. 7.5 and Fig. 7.6).

Another observation is that whereas the distribution of S^* in Fig. 7.6 is quite smooth, in Fig. 7.5 it is quite irregular. This can be explained by the following argument.

When $v = 0.7$, due to the presence of the powers of v in Eq. (7.2), S^* effectively becomes a sum of a small number of random variables. For example, already at time $t = 10$ with maximal value of X_t being $M = 10$, the actual cost value does not exceed $v^{t-1}M = 0.4$. Furthermore, in Markov reward model the X_ts are dependent random variables and may take the same values for successive values of t due to the step structure of costs for health states (see Fig. (7.2)). In the model with iid costs, $S^{(1)}$, $S^{(2)}$ and $S^{(3)}$ are mixtures of independent random variables. This explains the smoothness of $S^{(1)}$, $S^{(2)}$ and $S^{(3)}$ and the irregularity that we observe for S^*. When $v = 1.02$, S^* is a sum of a large number of random variables and, despite their dependence, we can observe a smooth curve.

We also remark that when $v = 0.7$ and the maximal cost value M equals 10, the maximal value of S in Eq. (7.2) is bounded by

$$\max S \leq M \sum_{j \geq 0} v^j = \frac{M}{1-v} = \frac{10}{1-0.7} \simeq 33. \tag{7.35}$$

For $v = 1.02$, S takes a wide spectrum of values, which is bounded by

$$\max S \leq M \sum_{j=0}^{\bar{L}} v^j = M \frac{1-v^{\bar{L}+1}}{1-v}, \tag{7.36}$$

where $\bar{L}$ is the maximal number of the years of life, which is determined analogously to Eq. (5.21) for the maximal service time in a pension plan. In our example, for an individual aged x $\bar{L}$ is about 75 years with the precision of 1%. Thus,

$$\max S \leq 10 \frac{1-1.02^{75+1}}{1-1.02} \simeq 1752. \tag{7.37}$$

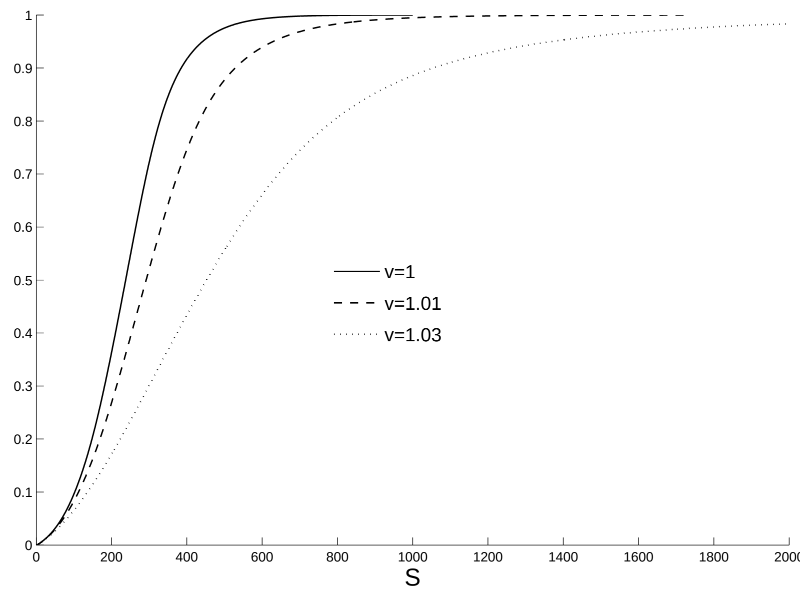

Figure 7.7: Markov reward model: distribution of S for different $v > 1$
W_i, $i \in \{A\}$ from Fig. 7.2, SW1911M

7.5 Sensitivity analysis

In practice, there are many health and non-health related phenomena that may occur and lead to significant changes in future profits and losses of a health care contract.

One of such phenomena is the change of the coefficient v due to market fluctuations. The impact of v on the distribution of S, computed using the Markov reward model for costs, is demonstrated in Fig. 7.7 for $v \geq 1$ and in Fig. 7.8 for $v < 1$. We give two different graphs because we need two different horizontal scales. Both figures show that by increasing v we increase the spread between the distributions, because, by increasing v, we increase the values S can take, which leads to the shift of the probability mass to higher values.

Another phenomenon, which becomes more important with new developments in medicine and the accompanying change in mortality, is that people survive longer in older ages and in bad health states. To estimate the financial impact of such longevity we perform a perturbation of matrix Λ with respect to mortality rates, similarly to Eq. (2.29); in the present case, we apply the change to the last m health states only. In Fig. 7.9 we present the distributions of S , computed using the Markov reward

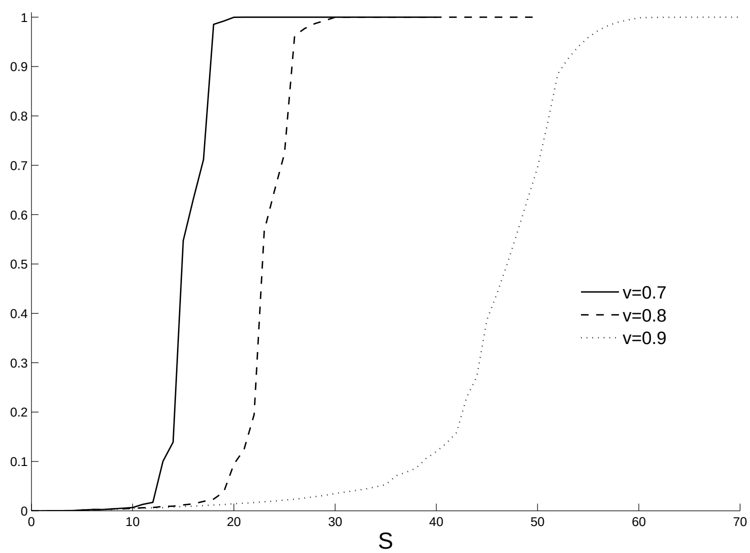

Figure 7.8: Markov reward model: distribution of S for different $v \leq 1$ W_i, $i \in \{A\}$ from Fig. 7.2, SW1911M

model for costs, taking $v = 1$ and $m = 30$ and different values of the perturbation coefficient ε. One observes from the figure that the decrease of mortality rates for the last 30 health states leads to the change of the distribution only for costs greater than 200 approximately. This indicates that the individuals who experience the change of mortality rates also spend more than 200. This amount may be compared to the expected cost until reaching state $n - m$. In our examples, the individuals have age 40 at the start of the contract, which implies that their expected health state is about 90 (see Eq. (4.12)). The average time spent in state i equals $1/|\Lambda_{ii}|$ and the health cost in state i is W_i. Thus, the expected cost accumulated from state 90 to state $n - m = 170$ is given by $\sum_{i=90}^{170} W_i/|\Lambda_{ii}| \simeq 209$, which confirms the observation.

7.6 Fluid queue approach

Our objective in this section is to compute the distribution of

$$S_t = \int_0^t v^u X_u du, \tag{7.38}$$

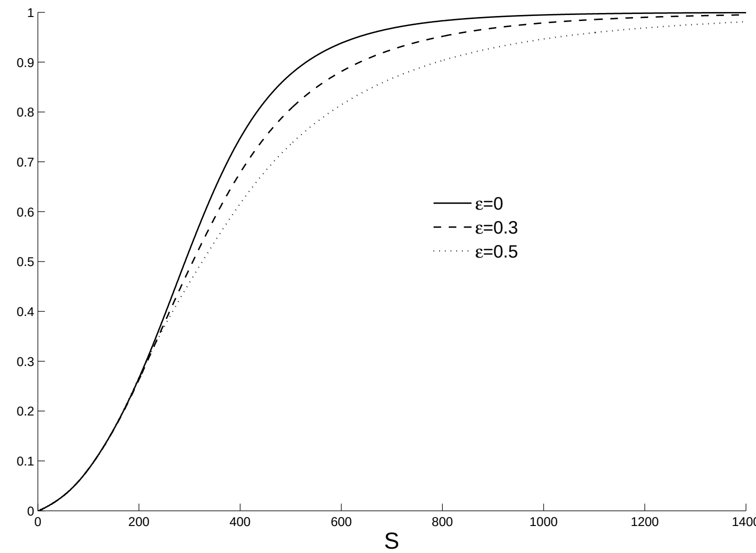

Figure 7.9: Markov reward model: distribution of S with mortality rates decreased for the last 30 states by ε $v = 1$, W_i, $i \in \{A\}$ from Fig. 7.2, SW1911M

where t is a fixed time horizon, X_u is a health cost rate at time u, v is a constant.

The *fluid queue* is a continuous version of the Markov reward model, to be used to describe the dynamic of S_t in continuous time. According to Section 1.2, we define the fluid queue as a two dimensional Markov process $\{(S_t, \phi_t),\, t \in \mathbb{R}^+\}$, where:

- $S_t \in \mathbb{R}^+$ is the "level" at time t, the net present value of a health care contract accumulated over the interval $(0, t)$;

- ϕ_t is the health states process at time t, it can take values in a finite state space, $A \cup \{D\}$, and is controlled by the generator Q of the aging process, defined in Eq. (2.1).

During intervals of time when ϕ_t is constant and equal to i, the level S_t varies at the time-dependent rate $r_i v^t$, which, in our case, is positive if the individual is still alive and in one of the health states in A, and equal to zero for the state $\{D\}$. The evolution of the net present value can thus be expressed by the following equation:

$$dS_t/dt = r_{\phi_t} v^t. \tag{7.39}$$

Here, r_{ϕ_t} plays the role of health care cost rate X_t at time t.

In order to obtain the distribution of S_t, we condition on the health state at time 0

$$\mathrm{P}\left[\, S_t \le z \,\right] = \underline{\tau}\, \underline{H}_t^{\mathrm{T}}(z), \quad \text{with} \quad H_{t,i}(z) = \mathrm{P}\left[\, S_t \le z \,|\, \phi_0 = i \,\right], \tag{7.40}$$

and we focus on the conditional probability given ϕ_0.

We start our discussion with the assumption that Q is a two-diagonal generator like the one presented in Section 2.1 for the PH-aging model. We notice that, due to the two-diagonal structure of Q presented in Fig. 2.1, if at time 0 the individual is in the health state i out of n possible active states and one state for death, then by time t the individual can make at most $n+1-i$ health state transitions. This argument allows us to express $\underline{H}_t(z)$ as follows in terms of the joint distribution of the level and the health state at time t given the health state at time 0

$$H_{t,i}(z) = \sum_{j=0}^{n-i+1} F_{i,t,j}(z), \tag{7.41}$$

where

$$F_{i,t,j}(z) = \mathrm{P}\left[\, S_t \le z, \phi_t = i+j \,|\, \phi_0 = i \,\right]. \tag{7.42}$$

This allows us to focus on the joint probabilities $F_{i,t,j}(z)$ and express them recursively. Denote by $\delta_i(u,t)$ the *increment of the level* from time u to time t if the health state remains continuously in state i. It is determined by

$$\delta_i(u,t) = \int_u^t r_i v^s ds = r_i(v^t - v^u)/\log v. \tag{7.43}$$

Theorem 7.6.1 *If Q is a two-diagonal generator, the joint distribution functions $F_{i,t,k}(z)$, $z \ge 0$ are recursively given by*

$$F_{i,t,k}(z) = \int_0^t F_{i,u,k-1}\left(z - \delta_{i+k}(u,t)\right) Q_{i+k-1,i+k} (e^{Q(t-u)})_{i+k,i+k}\, du, \tag{7.44}$$

for $k = 1, ..., n-i+1$. The initial condition corresponds to $k=0$ and is given by

$$F_{i,t,0}(z) = (e^{Qt})_{i,i} \mathbb{1}_{\{\delta_i(0,t) \le z\}}. \tag{7.45}$$

Proof. The proof follows from a simple renewal argument, which is illustrated in Fig. 7.10. If after having made k health state transitions by time t, given health state i at time 0, the level is at most z, it means that there was some time u in the past, at which the kth transition took place. The increment between time u to time

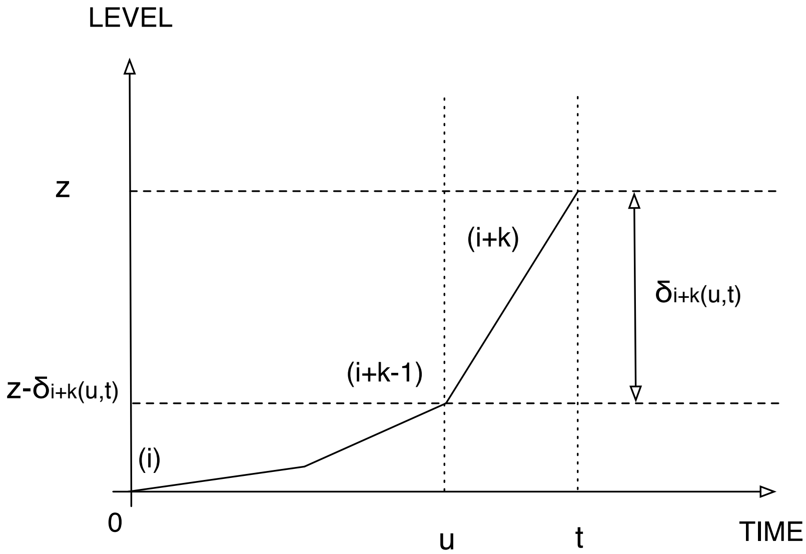

Figure 7.10: Net present value S_θ, $0 \leq \theta \leq t$. Illustration

t if the health state is $i+k$ equals $\delta_{i+k}(u,t)$, which signifies that at time u the level should have been at most equal to $z - \delta_{i+k}(u,t)$. ■

We derive the algorithm to determine $F_{i,t,k}(z)$ from the expressions given in the following theorem by solving numerically the integrals in Eq. (7.44). The distributions of S_t within one year period, calculated numerically for integer z with a three months interval, are presented in Fig. 7.11. Obviously, S_0 equals zero. As t increases, the fluid is allowed to make more transitions, and the distribution of S_t becomes smoother, in addition to being shifted to the right.

Let us now assume that the generator Q has no special structure. The two-diagonal structure allowed us to count the number of transitions in the interval $(0,t)$ and connect them to the state at time t so that $H_{t,i}(z)$ is the sum of the joint probabilities $F_{i,t,k}(z)$, as shown in Eq. (7.41). If the generator Q has an arbitrary structure, then the number of transitions in the interval $(0,t)$ is unbounded and is no longer directly connected to the state at time t. Thus, we need a slightly different approach. Denote

$$F_{i,t,j,k}(z) = \mathrm{P}\left[\, S_t \leq z, \phi_t = j, n_t = k \,|\, \phi_0 = i \,\right],$$

where n_t is the number of transitions in the interval $(0,t)$. Then,

$$H_{t,i}(z) = \sum_{j \in A \cup \{D\}} \sum_{k=0}^{\infty} F_{i,t,j,k}(z), \tag{7.46}$$

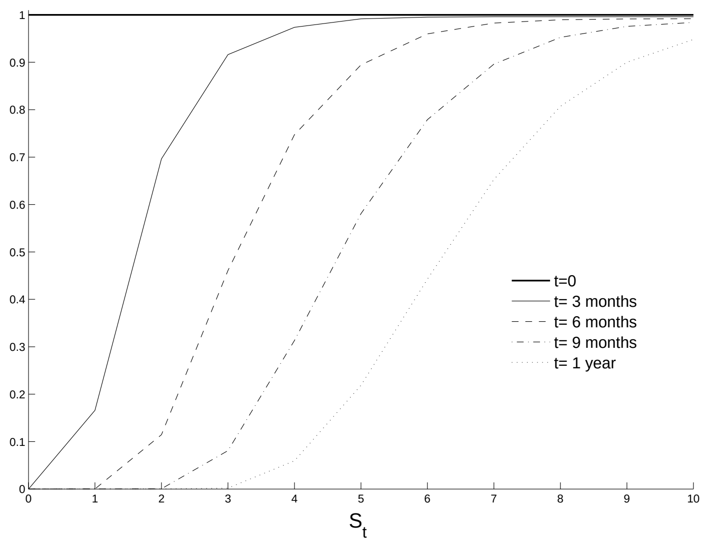

Figure 7.11: Distribution of S_t, $v = 0.8$
$t = 0$, 0.25, 0.5, 0.75 and 1 year, r_i, $i \in \{A\}$ from Fig. 7.2, SW1911M

and an algorithm to determine $F_{i,t,j,k}(z)$ may be derived from the following theorem.

Theorem 7.6.2 *The joint distribution functions $F_{i,t,j,k}(z)$, $z \geq 0$ are recursively given by*

$$F_{i,t,j,k}(z) = \int_0^t \sum_{\tilde{j} \in A \cup \{D\}} F_{i,u,\tilde{j},k-1}\left(z - \delta_j(u,t)\right) Q_{\tilde{j},j} (e^{Q(t-u)})_{j,j}\, du, \tag{7.47}$$

for $k = 1, ..., \infty$. The initial condition corresponds to $k = 0$ and is given by

$$F_{i,t,j,0}(z) = (e^{Qt})_{i,i} \mathbb{1}_{\{\delta_i(0,t) \leq z\}} \mathbb{1}_{\{i=j\}}. \tag{7.48}$$

Here, Q is a generator of any structure.

Proof. We apply the same type of argument as in Theorem 7.6.1 to determine $F_{i,t,j,k}(z)$. The only difference is that at time u the individual is allowed to be in any state, and therefore we need to sum over all possible states inside the integral (7.47).
■

Chapter 8

Stochastic extensions

The chapter is devoted to the development and analysis of some stochastic extensions to the models introduced in Chapter 7.

We start the chapter with our motivation to develop such models, which we present in Section 8.1. In Section 8.2, we work in discrete time and develop the stochastic analogue of the Markov reward model for costs, defined in Section 7.3.

In continuous time, we extend the fluid queue approach introduced in Section 7.6 by assuming the health care cost rate to be random for a given health state. This automatically implies that the *increment of the level* $\delta_i(u, t)$, defined in Eq. (7.43), is a random variable, and we need to describe this random variable. It appears that the randomness in $\delta_i(u, t)$ can be introduced in several ways. In Section 8.3 and Section 8.4, we present two fluid models with different assumptions for $\delta_i(u, t)$.

8.1 Motivation

We explain our motivation with the help of Fig. 8.1. The graph shows the simulation of health care costs in the future life years of an individual aged 40.

The dots represent a trajectory of independent identically distributed health care costs; the crosses represent health care costs that follow the Markov reward model and, thus, take constant values for a given health state. The distribution of the net present value of health care policies under these two assumptions for costs was

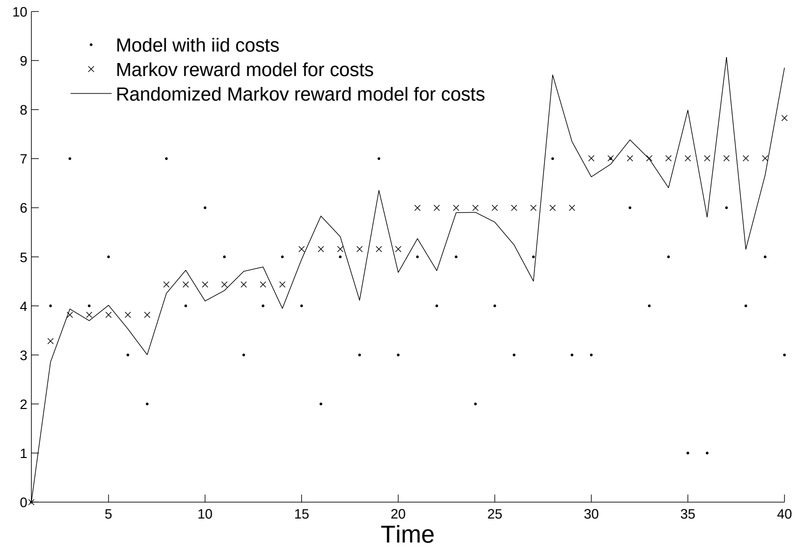

Figure 8.1: Health care costs per year. Illustration

obtained in Section 7.2, and in Section 7.3 (in discrete time) and 7.6 (in continuous time).

One sees from the figure that, depending on a particular type of health care, both models may not have a perfect physical interpretation. In the iid model the costs are a bit too random, even if the health care concerns only exogenous health damages; furthermore, for general type of health care it is not very likely to have high costs in younger ages and lower costs in older ages. It is not very likely either to have such a deterministic step structure for costs, as shown by the crosses.

One natural extension is to assume that the costs do generally increase with the health state, and do so with an element of randomness, as shown by the solid line.

8.2 Randomized Markov reward model

We define a *randomized Markov reward model* $(A \cup \{D\}, \underline{\mathcal{W}}, P)$ (below, "randomized MRM") as an extension of the Markov reward model for costs, defined in Section 7.3 and given by $(A \cup \{D\}, \underline{W}, P)$. In Section 7.3, it is assumed that, if the health state is i, $i \in A \cup \{D\}$, then the cost X_t in Eq. (7.2) is a known constant W_i. Here, we suppose that $\mathcal{W}_i$ is a discrete random variable: it is defined on a given set $\mathcal{C} = \{c_1, ..., c_M\}$ of non-negative values, if $i \in A$, with a distribution which may depend on i, and it is equal to 0 with probability one if $i = \{D\}$. The distribution function and the density of $\mathcal{W}_i$ are

$$F^{(i)}(k) = \mathrm{P}[\,\mathcal{W}_i \le k\,] \quad \text{and} \quad f^{(i)}(k) = \mathrm{P}[\,\mathcal{W}_i = k\,], \quad i \in A \cup \{D\},\ k \ge 0. \tag{8.1}$$

Theorem 8.2.1 *Suppose that $\{X_t\}$ follows the randomized Markov Reward model defined by the triplet $(A \cup \{D\}, \underline{\mathcal{W}}, P)$. The conditional distribution $\underline{H}(k)$ of the net present value S, defined by Eq. (7.2), is*

$$H_i(k) = y_i F^{(i)}(k) + (e^{\Lambda})_{(i,\cdot)} \sum_{j=1}^{M} f^{(i)}(c_j) \underline{H}^T\left(\frac{k - c_j}{v}\right), \;\; k \ge 0, \;\; i = 1, .., n, \tag{8.2}$$

where $\underline{y}$ is given by Eq. (7.6).

Proof. Similarly to the proofs of Theorem 7.2.1 and Theorem 7.3.1, we obtain

$$\begin{aligned} H_i(k) =& \mathrm{P}\,[\,[L] = 1 \,|\, \phi_0 = i\,]\ \mathrm{P}[X_1 \le k \,|\, \phi_0 = i\,] + \\ & \sum_{s \in \{A\}} \mathrm{P}[\,\phi_1 = s \,|\, \phi_0 = i\,]\,\mathrm{P}[\,X_1 + v\tilde{S} \le k \,|\, \phi_1 = s, \phi_0 = i\,] \\ =& y_i F^{(i)}(k) + \sum_{s \in \{A\}} (e^{\Lambda})_{(i,s)} \sum_{j=1}^{M} f^{(i)}(c_j) H_s\left(\frac{k - c_j}{v}\right) \\ =& y_i F^{(i)}(k) + \sum_{j=1}^{M} f^{(i)}(c_j) \sum_{s \in \{A\}} (e^{\Lambda})_{(i,s)} H_s\left(\frac{k - c_j}{v}\right) \\ =& y_i F^{(i)}(k) + (e^{\Lambda})_{(i,\cdot)} \sum_{j=1}^{M} f^{(i)}(c_j) \underline{H}^{\mathrm{T}}\left(\frac{k - c_j}{v}\right). \end{aligned} \tag{8.3}$$

■

One may notice that Eq. (8.2) is very similar to Eq. (7.7); the only difference is that the probability mass function $f^{(i)}(c_j)$ and the cumulative distribution function $F^{(i)}(k)$ have index (i), which indicates that the distribution of cost X_1 depends on health state i. The following lemma determines the expectation of S and goes without proof as it immediately results from Lemma 7.3.2.

Lemma 8.2.2 *The expectation of S in the randomized Markov reward model $(A \cup \{D\}, \underline{\mathcal{W}}, P)$ is*

$$E[S] = [\underline{\tau}\ 0](I - vP)^{-1} E[\underline{\mathcal{W}}^T], \tag{8.4}$$

where the matrix P is defined by Eq. (7.12).

8.3 Fluid model with geometric cost rates

In Section 7.6, we assume that the rate r_{ϕ_t} is the constant r_i as long as ϕ_t remains equal to i. Here, we assume that it is a random variable, which takes only non-negative values. Specifically, we shall assume in this section that r_{ϕ_t}, given $\phi_t = i$, evolves like a geometric Brownian motion $R_t^{(i)}$ of the form given in Appendix A.3, with parameters μ_i, σ_i and $R_0^{(i)}$. Thus, the increment of the level from time u to time t for state i is given by

$$\delta_i(u,t) = \int_u^t v^s R_s^{(i)} ds, \quad 0 \le u \le t, \tag{8.5}$$

and we write $v = e^{-\alpha}$, where α is the *force of interest*. To give a better feeling for the behavior of costs in such model we give an illustration in Fig. 8.2. In this simple example we change state i to state $i+1$ every 30 years; the drift μ_i increases with the state and the diffusion coefficient σ_i is assumed to be independent of i. The jumps that we observe at the moments of changing the state are due to the fact that the stochastic processes $R_t^{(i)}$, for all i, start evolving at $t = 0$, and are independent of each other.

We establish in the following lemma a connection between two stochastic processes, $\delta_i(u,t)$ and $\delta_i(0, t-u)$.

Lemma 8.3.1 *If $\delta_i(u,t)$ is defined by Eq.* (8.5)*, then*

$$\delta_i(u,t) \stackrel{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} e^{-\alpha u} \delta_i(0, t-u),$$

where $\tilde{R}_u^{(i)}$ is a geometric Brownian motion with parameters μ_i, σ_i and $R_0^{(i)}$, and independent on $R_u^{(i)}$.

Proof. According to Eq. (A.10),

$$R_t^{(i)} = R_0^{(i)} e^{(\mu_i - \sigma_i^2/2)t + \sigma_i W_t}, \tag{8.6}$$

where W_t is a standard Brownian motion. For the simplicity of notations, denote

$$\bar{\mu}_i = \mu_i - \sigma_i^2/2, \quad \bar{\sigma}_i = \sigma_i. \tag{8.7}$$

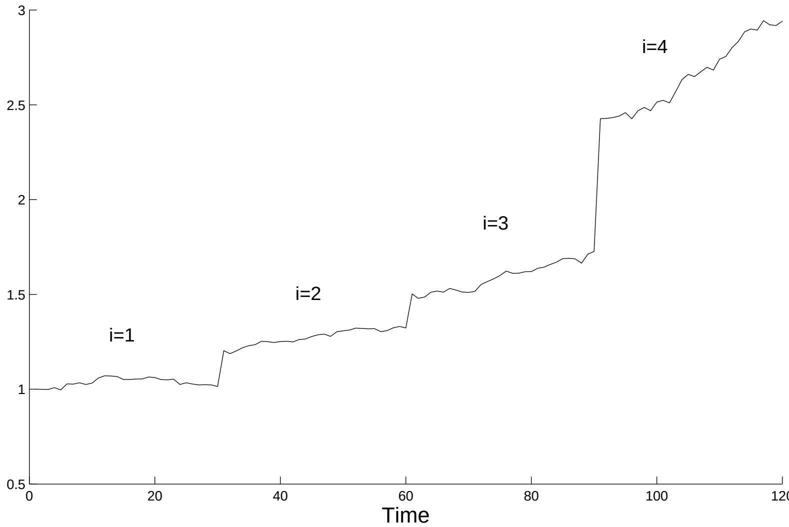

Figure 8.2: Health care costs: state-dependent geometric Brownian motion
Parameters: $\mu_1 = 5 \cdot 10^{-4},\ \mu_i = \mu_{i-1} + 3 \cdot 10^{-3}, \quad i = 2-4$
$\sigma_i = 1.5 \cdot 10^{-2},\ R_0^{(i)} = 1, \quad i = 1-4.$

We may rewrite Eq. (8.6) as

$$\begin{aligned} R_t^{(i)} &= R_0^{(i)} e^{\bar{\mu}_i t + \bar{\sigma}_i W_t} \\ &= R_0^{(i)} e^{\bar{\mu}_i (t-u) + \bar{\mu}_i u + \bar{\sigma}_i (W_t - W_{t-u}) + \bar{\sigma}_i W_{t-u}} \\ &= R_0^{(i)} e^{\bar{\mu}_i (t-u) + \bar{\sigma}_i W_{t-u}} e^{\bar{\mu}_i u + \bar{\sigma}_i (W_t - W_{t-u})}. \end{aligned}$$

Since W_t has stationary and independent increments,

$$\begin{aligned} R_t^{(i)} &\stackrel{d}{=} R_0^{(i)} e^{\bar{\mu}_i (t-u) + \bar{\sigma}_i W_{t-u}} e^{\bar{\mu}_i u + \bar{\sigma}_i \widetilde{W}_u} \\ &= \frac{1}{R_0^{(i)}} R_{t-u}^{(i)} \tilde{R}_u^{(i)}. \end{aligned} \tag{8.8}$$

Here, $\widetilde{W}_u$ is a standard Brownian motion, independent on W_u, so that $\tilde{R}_u^{(i)}$ is independent of $R_{t-u}^{(i)}$.

By definition (8.5) of $\delta_i(u,t)$, and by Eq. (8.8), we obtain

$$\begin{aligned}\delta_i(u,t) &= \int_u^t e^{-\alpha s} R_s^{(i)} ds \overset{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} \int_u^t e^{-\alpha s} R_{s-u}^{(i)} ds \\ &\overset{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} e^{-\alpha u} \int_0^{t-u} e^{-\alpha s} R_s^{(i)} ds \\ &\overset{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} e^{-\alpha u} \delta_i(0, t-u).\end{aligned}$$

■

Denote

$$Y_\alpha^{(i)}(t) = \int_0^t e^{(\bar{\mu}_i - \alpha)s + \bar{\sigma}_i W_s} ds, \tag{8.9}$$

where t is a fixed time horizon. Integrals of this type are known and widely used in mathematical finance. The distribution of $Y_\alpha^{(i)}(t)$ was explicitly obtained in 1992 by Marc Yor [66] and studied further in Schröder [53]. Since the distribution does not have a very simple form, in our computations below we find it convenient to denote by $g^{(i)}(x,t;\alpha)$ and by $G^{(i)}(x,t;\alpha)$ the probability density and the cumulative distribution function of $Y_\alpha^{(i)}(t)$, respectively.

Denote the cumulative distribution function of a *lognormal random variable* $X = x_0 e^{\mu + \sigma Z}$, where x_0, μ and σ are parameters, and Z is a standard normal random variable, by $l(x;\mu,\sigma^2,x_0)$. According to Appendix A.2, it is given by

$$l(x;\mu,\sigma^2,x_0) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - (\ln x_0 + \mu))^2}{2\sigma^2}}. \tag{8.10}$$

Lemma 8.3.2 *The cumulative distribution function of* $\delta_i(u,t)$, $0 \le u \le t$ *defined by Eq.* (8.5) *with* $R_0^{(i)} > 0$ *is*

$$\mathrm{P}[\, \delta_i(u,t) \le z\,] = G^{(i)}(z/R_0^{(i)}, t; \alpha), \quad u = 0, \tag{8.11}$$

and for $u > 0$ *we have*

$$\mathrm{P}[\, \delta_i(u,t) \le z\,] = \int_0^\infty l(s;(\bar{\mu}_i - \alpha)u, \bar{\sigma}_i^2 u, R_0^{(i)}) G^{(i)}\,(z/s, t-u;\alpha)\, ds, \tag{8.12}$$

The parameters $\bar{\mu}_i$ *and* $\bar{\sigma}_i$ *are defined in Eq.* (8.7).

Proof. To prove Eq. (8.11) we express $\delta_i(0,t)$ in terms of $Y_\alpha^{(i)}(t)$

$$\delta_i(0,t) = \int_0^t e^{-\alpha s} R_s^{(i)} ds = \int_0^t e^{-\alpha s} R_0^{(i)} e^{\bar{\mu}_i s + \bar{\sigma}_i W_s} ds$$
$$= R_0^{(i)} \int_0^t e^{(\bar{\mu}_i - \alpha)s + \bar{\sigma}_i W_s} ds \stackrel{d}{=} R_0^{(i)} Y_\alpha^{(i)}(t).$$

Therefore,

$$\mathrm{P}[\,\delta_i(0,t) \le z\,] = \mathrm{P}\left[\,R_0^{(i)} Y_\alpha^{(i)}(t) \le z\,\right]$$
$$= \mathrm{P}\left[\,Y_\alpha^{(i)}(t) \le z/R_0^{(i)}\,\right] = G^{(t)}(z/R_0^{(i)}, t; \alpha).$$

To prove Eq. (8.12) we use Lemma 8.3.1 and Eq. (8.11) to express $\delta_i(u,t)$, $u > 0$ in terms of $Y_\alpha^{(i)}(t-u)$

$$\begin{aligned} \delta_i(u,t) &\stackrel{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} \delta_i(0, t-u) e^{-\alpha u} \\ &\stackrel{d}{=} \frac{1}{R_0^{(i)}} \tilde{R}_u^{(i)} e^{-\alpha u} R_0^{(i)} Y_\alpha^{(i)}(t-u) \\ &\stackrel{d}{=} \tilde{R}_u^{(i)} e^{-\alpha u} Y_\alpha^{(i)}(t-u). \end{aligned} \tag{8.13}$$

By Eq. (8.6), Eq. (8.13) and the definition of *convolution* operation (see Appendix A.4),

$$\begin{aligned} \mathrm{P}[\,\delta_i(u,t) \le z\,] &= \mathrm{P}\left[\,\tilde{R}_u^{(i)} e^{-\alpha u} Y_\alpha^{(i)}(t-u) \le z\,\right] \\ &= \mathrm{P}\left[\tilde{R}_0^{(i)} e^{(\bar{\mu}_i - \alpha)u + \bar{\sigma}_i W_u} Y_\alpha^{(i)}(t-u) \le z\right] \\ &= \int_0^\infty l(s; (\bar{\mu}_i - \alpha)u, \bar{\sigma}_i^2 u, R_0^{(i)}) G^{(i)}\left(z/s, t-u; \alpha\right) ds, \end{aligned}$$

because $\tilde{R}_0^{(i)} e^{(\bar{\mu}_i - \alpha)u + \bar{\sigma}_i W_u}$ is a *lognormal random variable* with parameters $x_0 = \tilde{R}_0^{(i)} = R_0^{(i)}$, $\mu = (\bar{\mu}_i - \alpha)u$ and $\sigma = \bar{\sigma}_i \sqrt{u}$. ■

If Q is a two-diagonal generator like in PH-aging model, one derives the algorithm to determine $F_{i,t,k}(z) = \mathrm{P}\left[\,S_t \le z, \phi_t = i+k \,|\, \phi_0 = i\,\right]$ from the following theorem.

Theorem 8.3.3 *Assume that $\{r_t\}$ is a geometric Brownian motion with parameters that are state-dependent and $\delta_i(u,t)$, $0 \le u \le t$, is defined by Eq. (8.5). If Q is a* *two-diagonal generator, the joint distribution functions* $F_{i,t,k}(z)$, $z \geq 0$ *are recursively given by*

$$F_{i,t,k}(z) = \int_0^t \int_0^z F_{i,u,k-1}(z-y)\, \tilde{g}_\alpha^{(i+k)}(y;u,t) Q_{i+k-1,i+k} (e^{Q(t-u)})_{i+k,i+k}\, dydu \quad (8.14)$$

for $k = 1, ..., n-i+1$. *The initial condition corresponds to* $k = 0$ *and is given by*

$$F_{i,t,0}(z) = (e^{Qt})_{i,i} G^{(i)}(z/R_0^{(i)}, t; \alpha). \quad (8.15)$$

Here,

$$\tilde{g}_\alpha^{(i)}(y;u,t) = \int_0^\infty \frac{1}{s} l(s; (\bar{\mu}_i - \alpha)u, \bar{\sigma}_i^2 u, R_0^{(i)}) g^{(i)}(y/s, t-u; \alpha)\, ds, \quad (8.16)$$

where $\bar{\mu}_i$ *and* $\bar{\sigma}_i$ *are defined in Eq.* (8.7).

Proof. In general, the proof repeats the one of Theorem 7.6.1. Here, to prove Eq. (8.15), we only need to demonstrate that $G^{(i)}(z/R_0^{(i)}, t; \alpha)$ is the cumulative distribution function of $\delta_i(0,t)$. This immediately follows from Lemma 8.3.2 for $u = 0$. To prove Eq. (8.14), we only need to show that the probability density function of $\delta_i(u,t)$ is $\tilde{g}_\alpha^{(i)}(y;u,t)$, which is given by Eq. (8.16). This is a direct result from the differentiation of Eq. (8.12) in Lemma 8.3.2 with respect to z. This completes the proof. ■

If Q has no special structure, we use the same argument as in Theorem 8.3.3 to extend Theorem 7.6.2 as follows.

Theorem 8.3.4 *Assume that that* $\{r_t\}$ *is a geometric Brownian motion with state-dependent parameters and* $\delta_i(u,t)$, $0 \leq u \leq t$, *is defined by Eq.* (8.5). *Assume also that* $\tilde{g}_\alpha^{(i)}(y;u,t)$ *is given by Eq.* (8.16). *The joint distribution functions* $F_{i,t,j,k}(z)$, $z \geq 0$ *are recursively given by*

$$F_{i,t,j,k}(z) = \int_0^t \sum_{\tilde{j} \in A \cup \{D\}} \int_0^z F_{i,u,\tilde{j},k-1}(z-y)\, \tilde{g}_\alpha^{(j)}(y;u,t) Q_{\tilde{j},j} (e^{Q(t-u)})_{j,j}\, dudy, \quad (8.17)$$

for $k = 1, ..., \infty$. *The initial condition corresponds to* $k = 0$ *and is given by*

$$F_{i,t,j,0}(z) = (e^{Qt})_{i,i} G^{(i)}(z/R_0^{(i)}, t; \alpha) \mathbb{1}_{\{i=j\}}. \quad (8.18)$$

Here, Q *is a generator of any structure.*

8.4 Fluid model with Brownian increments

We assume that X_u in Eq. (7.38) can potentially take negative values. This assumption leads to the fact that, for a given health state i, the increment $\delta_i(u,t)$ of the level can take negative values. Thus, we may represent $\delta_i(u,t)$ as a discounted increment from time u to time t of stochastic process $B_s^{(i)}$,

$$\delta_i(u,t) = \int_u^t v^s dB_s^{(i)}, \quad 0 \le u \le t, \tag{8.19}$$

where $B_s^{(i)}$ is a *Brownian motion*, defined in Appendix A.3, with parameters that depend on state i: drift μ_i and variance σ_i; $v = e^{-\alpha}$, where α is the force of interest.

We give an example of the cost behavior in Fig. 8.3. We choose 1 as a time discretization step for $B_t^{(i)}$, which has led to

$$\begin{aligned} B_{t+1}^{(i)} - B_t^{(i)} &= \mu_i(t+1) - \mu_i t + \sigma_i W(t+1) - \sigma_i W(t) \\ &\stackrel{d}{=} \mu_i + \sigma_i W(1), \end{aligned}$$

where $W(t)$ is a standard Brownian motion, defined in Appendix A.3. Thus, in state i the costs fluctuate around constant μ_i with variance σ_i^2. In Fig. 8.3 we change state i to state $i+1$ after a random period of time given by $1/|\Lambda(i,i)|$, which is the average sojourn time in state i in the PH-aging model. The drift μ_i is chosen to increase with the state, and the variance σ_i^2 is chosen to be state-independent.

Similarly to Section 8.3, $\delta_i(u,t)$ is a stochastic process and the connection between $\delta_i(u,t)$ and $\delta_i(0,t-u)$ is given in the following lemma.

Lemma 8.4.1 *If $\delta_i(u,t)$ is defined by Eq.* (8.19)*, then $\delta_i(u,t) \stackrel{d}{=} \delta_i(0,t-u)e^{-\alpha u}$.*

Proof. By definition (8.19) of $\delta_i(u,t)$,

$$\begin{aligned} \delta_i(u,t) = \delta_i(u,u+t-u) &= \int_u^{u+t-u} e^{-\alpha s} dB_s^{(i)} && (8.20) \\ &= \int_0^{t-u} e^{-\alpha(s+u)} dB_{s+u}^{(i)} = e^{-\alpha u} \int_0^{t-u} e^{-\alpha s} dB_{s+u}^{(i)}. && (8.21) \end{aligned}$$

The process $B_t^{(i)}$ has independent and stationary increments, therefore $dB_s^{(i)}$ has the same distribution as $dB_{s+u}^{(i)}$, and we have the statement of the lemma. ∎

One readily adapts Theorems 7.6.1 and 7.6.2 to the present model as shown below, firstly for a two-diagonal generator Q, secondly for the general case.

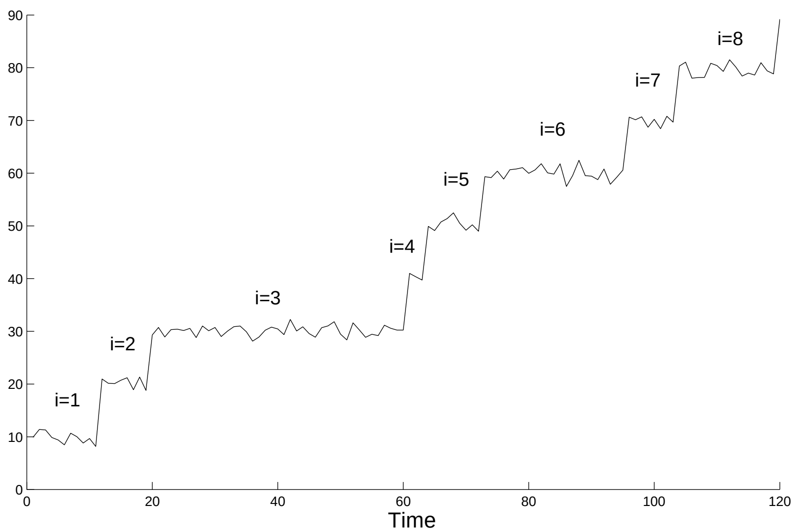

Figure 8.3: Health care costs: state-dependent Brownian increments
Parameters: $\mu_i = 10i,\ \sigma_i = 5,\ \ i = 1-8$, SW1911M

Theorem 8.4.2 *Assume that the increment $\delta_i(u,t)$ is given by Eq. (8.19) and Q is a two-diagonal generator. The joint distribution functions $F_{i,t,k}(z)$, $z \in \mathbb{R}$ are recursively given by*

$$F_{i,t,k}(z) = \int_0^t \int_{-\infty}^{\infty} F_{i,u,k-1}(z-y)\,\phi(y;\mu_{i+k}(u,t),\sigma^2_{i+k}(u,t))Q_{i+k-1,i+k}(e^{Q(t-u)})_{i+k,i+k}\,dydu \tag{8.22}$$

for $k = 1,...,n-i+1$. The initial condition corresponds to $k = 0$ and is given by

$$F_{i,t,0}(z) = (e^{Qt})_{i,i}\Phi(z;\mu_i(0,t),\sigma_i^2(0,t)). \tag{8.23}$$

Here, functions $\Phi(x;\mu,\sigma^2)$ and $\phi(x;\mu,\sigma^2)$, respectively, are the cumulative distribution and the probability density of the normal distribution with parameters μ and σ^2; $\mu_i(u,t)$ is the expected value of $\delta_i(u,t)$,

$$\mu_i(u,t) = e^{-\alpha u}\frac{\mu_i}{\alpha}(1-e^{-\alpha(t-u)}), \tag{8.24}$$

and $\sigma_i^2(u,t)$ is the variance of $\delta_i(u,t)$,

$$\sigma_i^2(u,t) = e^{-2\alpha u}\frac{\sigma_i^2}{2\alpha}(1-e^{-2\alpha(t-u)}). \tag{8.25}$$

Proof. In general, the proof repeats the one of Theorem 7.6.1. Here, we only need to prove that $\delta_i(u,t)$ is a normal random variable with parameters $\mu_i(u,t)$ and $\sigma_i^2(u,t)$, given by Eq. (8.24) and Eq. (8.25), respectively.

The Brownian motion $B_t^{(i)}$ is the solution of the stochastic differential equation (see, for example, Lin [39])

$$dB_t^{(i)} = \mu_i dt + \sigma_i dW_t, \tag{8.26}$$

where W_t is a standard Brownian motion. Thus, from Eq. (8.19) for $u = 0$ we obtain

$$d\delta_i(0,t) = e^{-\alpha t}\mu_i dt + e^{-\alpha t}\sigma_i dW_t, \tag{8.27}$$

From Example 5.6 in Lin [39] it follows immediately that

$$\delta_i(0,t) \sim \mathbb{N}\left(\int_0^t e^{-\alpha s}\mu_i ds, \int_0^t e^{-2\alpha s}\sigma_i^2 ds\right) \tag{8.28}$$

By computing the integrals in (8.28), we obtain

$$\delta_i(0,t) \sim \mathbb{N}\left(\frac{\mu_i}{\alpha}(1-e^{-\alpha t}), \frac{\sigma_i^2}{2\alpha}(1-e^{-2\alpha t})\right). \tag{8.29}$$

From (8.29) and Lemma 8.4.1 it follows that

$$\delta_i(u,t) \sim \mathbb{N}\left(e^{-\alpha u}\frac{\mu_i}{\alpha}(1-e^{-\alpha(t-u)}), e^{-2\alpha u}\frac{\sigma_i^2}{2\alpha}(1-e^{-2\alpha(t-u)})\right). \tag{8.30}$$

Therefore,

$$\delta_i(u,t) \sim \mathbb{N}\left(\mu_i(u,t), \sigma_i^2(u,t)\right). \tag{8.31}$$

■

If Q has no special structure, we use the same argument as in Theorem 8.4.2 to extend Theorem 7.6.2 as follows.

Theorem 8.4.3 *Assume that the increment $\delta_i(u,t)$ is given by Eq. (8.19). The joint distribution functions $F_{i,t,j,k}(z)$, $z \in \mathbb{R}$ are recursively given by*

$$F_{i,t,j,k}(z) = \\ \int_0^t \sum_{\tilde{j}\in A\cup\{D\}} \int_{-\infty}^{\infty} F_{i,u,\tilde{j},k-1}(z-y)\,\phi(y;\mu_j(u,t),\sigma_j^2(u,t))Q_{\tilde{j},j}(e^{Q(t-u)})_{j,j}\,dudy \tag{8.32}$$

for $k = 1, ..., \infty$. The initial condition corresponds to $k = 0$ and is given by

$$F_{i,t,j,0}(z) = (e^{Qt})_{i,i}\Phi(z;\mu_i(0,t),\sigma_i^2(0,t))\mathbb{1}_{\{i=j\}}. \tag{8.33}$$

Here, Q is a generator of any structure.

We remark that the constructed fluid model with Markov modulated Brownian increments is a continuous time version of the randomized MRM from Section 8.2. The parametrization procedure is essentially the same as for the randomized MRM: one needs to choose the expected costs μ_i for each state i and the level of uncertainty around this expectation. For example, μ_i might be chosen to be the same as for Markov reward model introduced in Section 7.3, and the variance might be chosen as a constant or proportional to μ_i.

In comparison with the fluid model with geometric cost rates introduced in Section 8.3, the fluid model with Markov modulated Brownian increments provides a more efficient algorithm to compute the distribution of S_t. The reason is that in Theorems 8.4.2 and 8.4.3 the random variable $\delta_i(u,t)$ has a normal distribution; in Theorems 8.3.3 and 8.3.4 the distribution of $\delta_i(u,t)$ depends on the distribution of $Y_\alpha^{(i)}(t-u)$ (see Eq. (8.13)), which is not easy to compute.

Another difference between the two fluid models is that the model with Markov modulated Brownian increments allows the costs X_u to take negative values. If X_u can only be positive, one needs to adapt the model. Clearly, if the average cost is around 500$ and the variance is 10$, the probability mass related to negative values of X_u is very small, and one may adopt the approximation. If one desires a more precise model, one may use a *subordinator* as a stochastic process for $B_s^{(i)}$ in Eq. (8.19). This would guarantee $dB_s^{(i)} \geq 0$ and, thus, the positivity of the costs. One example of such subordinator is a *gamma process* defined in Appendix A.3. Due to the discontinuity of sample paths (see Applebaum [2]), the replacement of Brownian motion $B_s^{(i)}$ by a gamma process leads to more complex calculations, which are not presented in this thesis, and which make a part of our future projects.

Conclusion and perspectives

In this thesis we have suggested several mathematical models for different life-linked insurances. The novelty of the models is that they use a phase-type distribution to describe the lifetime of an individual. This key assumption exhibits a number of nice properties of phase-type distributions, the most important of which being the connection that it provides between the age and the health state of an individual. The possibility to assign a reward structure to every health state naturally inspired us to compute the distribution of different lifetime dependent costs.

In pension insurance, we have computed distribution of the present value of future profits and losses of a defined benefit pension plan. Here, we extended the phase-type lifetime assumption to a multi-decrement case, which allowed us to solve the most challenging problem in profit-testing - to model the individuals in the pension plan. The problem is challenging as individuals tend not only to die and retire, but also to quit the pension plan for personal reasons (to "surrender") at random times.

In the part about health insurance, we have focused on the distribution of the net present value of health care costs. The phase-type approach gave us the opportunity to compute the distribution using a recursive procedures both in continuous and in discrete time. Furthermore, the flexibility of the phase-type approach allowed us to consider different assumptions for health costs, for example, a deterministic or a random cost for a given state of health. In addition, due to the phase-type lifetime assumption, it becomes convenient to employ fluid queue techniques to obtain the recursions in continuous time.

The phase-type lifetime assumption states that, while alive, an individual is traveling from one health state to another with certain speed until eventually he/she reaches the terminal ("dead") state. Changing the dynamic of this process has allowed us to model different population phenomena such as longevity and health change in a natural way. In both pension insurance and health care we applied the population modeling techniques to estimate the financial impact of health-related events.

During the construction of our models, we had to make a number of assumptions regarding the objects we examined. Evidently, one can go beyond these assumptions and consider more complex models to study different important events in insurance. For instance, one may introduce a correlation between surrenders and economical environment in the pension profit-test model, and so estimate the effect of mass surrendering of people in crisis times. For profitability estimation purposes, it would also be useful to introduce a detailed model for long-term interest rates. In the viewpoint of longevity problem, one may consider stochastic death rates in the phase-type lifetime model to better capture the uncertainty of the survival in the future. For instance, we find it interesting to employ the stochastic aging model not only to model correlated cohorts, but also to model the mortality in pension insurance and health care.

Evidently, a lot of work has to be done to bring these models to a practical use. In several places we gave indications of possible parametrization procedures; however, one needs to work with real data to make the models more practical and consider detailed case studies. One would also need to focus on fitting the model to specific examples.

We believe that the mathematical models we have constructed are general enough and are not tied to a particular field of insurance. The models we have developed in health care, for instance, allow to deal with any financial quantities that are life-dependent. In the next paragraphs we give some examples that can be considered as our future research directions.

Optimal consumption

In Part II and Part III we investigated questions related to the profitability of financial institutions such as pension funds and insurance companies. But what about households? Let us take the viewpoint of a retired individual with some savings. The question is how to spend efficiently these savings over the remaining period of life, and how to adjust the consumption to personal health and personal risk preferences. The motivation of this research came from Huang et al. [33], where the authors determine the optimal consumption over time in the classical lifecycle model first introduced by Yaari [65]. The authors suppose that the force of mortality obeys the law of Gompertz . In our approach, we assume that the lifetime of the individual is of phase-type, and below we present our first thoughts on the subject.

Huang et al. [33] considered the optimization problem

$$J = \max_{c(0)} E\left[\int_0^D e^{-\alpha t} u_\gamma(c(t)) 1_{\{t \le L\}} dt\right], \tag{8.34}$$

where $c(t)$ is the rate of consumption at time t, α is the force of interest, $u_\gamma(\cdot)$ is a *utility function* with *risk aversion coefficient* γ, D is a fixed time horizon and L is the remaining lifetime.

There are also budget constraints and boundary conditions, given by

$$\begin{cases} \dfrac{dF(t)}{dt} = rF(t) + \pi_0 - c(t), \\ F(0) = W \quad \text{and} \quad F(D) = 0, \end{cases} \tag{8.35}$$

where $F(t)$ is the capital trajectory at time t, W, $W > 0$ is the initial savings amount, π_0 is the constant income rate of the individual, r is a risk free interest rate.

In Huang et al. [33] the authors determine an explicit form of $c(t)$, assuming that the utility function is given by

$$\begin{cases} u(c) = \dfrac{c^{1-\gamma}}{1-\gamma}, & \gamma > 0,\, \gamma \neq 1, \\ u(c) = \ln c, & \gamma = 1. \end{cases} \tag{8.36}$$

This is the constant relative risk aversion (CRRA) type of utility functions (for more details, see P. P. Wakker [60]). The function $u(c)$ is strictly increasing, $\gamma \geq 0$ is the coefficient of risk aversion. If $\gamma = 0$, the individual is "risk-neutral", $\gamma \to \infty$ stands for "infinite risk aversion".

The optimal consumption rate is determined as

$$c_\gamma^*(t) = c_\gamma^*(0) e^{\frac{r-\alpha}{\gamma} t} (p_t)^{1/\gamma}, \tag{8.37}$$

where p_t is the t years survival probability of the individual. The initial optimal consumption rate $c_\gamma^*(0)$ is determined as follows. By substituting Eq. (8.37) into

(8.35) the authors in [33] arrive at the first-order ordinary differential equation for $F(t)$

$$\frac{dF(t)}{dt} - rF(t) - \pi_0 + c^*_\gamma(0)e^{\frac{r-\alpha}{\gamma}t}\left(p_t\right)^{1/\gamma} = 0,$$

the solution of which is

$$F(t) = e^{rt}\left(\pi_0 \int_0^t e^{-rs}ds - c^*_\gamma(0)\int_0^t e^{\frac{r-\alpha}{\gamma}s}\left(p_s\right)^{1/\gamma} e^{-rs}ds + W\right), \tag{8.38}$$

and $c^*_\gamma(0)$ is determined by the constraint $F(D) = 0$.

In our case, mortality for an individual aged x is driven by the PH-aging model and, therefore,

$$p_s = \underline{\tau}_x e^{\Lambda s}\mathbf{1}, \tag{8.39}$$

where $\underline{\tau}_x$ is the health state distribution at age x, defined by Eq. (2.3), transition rate matrix Λ is defined by Eq. (2.2). We combine (8.38) with (8.39) to obtain that

$$c^*_\gamma(0) = \frac{\pi_0/r(e^{rt}-1) + We^{rt}}{e^{rt}I_\gamma}, \tag{8.40}$$

where

$$I_\gamma = \int\limits_0^t \left(\underline{\tau}_x e^{((k-r)\gamma I + \Lambda)s}\mathbf{1}\right)^{1/\gamma} ds, \tag{8.41}$$

with $k = \dfrac{r-\alpha}{\gamma}$. If $\gamma = 1$, then

$$I_\gamma = \underline{\tau}_x\left((k-r)\gamma I + \Lambda\right)^{-1}\left(e^{((k-r)\gamma I+\Lambda)t} - I\right)\mathbf{1}. \tag{8.42}$$

If $\gamma \neq 1$, then it is more difficult to determine I_γ analytically, and we apply numerical methods.

We are interested in determining the sensitivity of $c^*(t)$ with respect to the risk aversion coefficient γ and with respect to the initial health state. In Fig. 8.4 we plot $c^*(t)$ for different values of γ and we make two observations. Firstly, we observe that the higher the γ, the closer the graphs to the flatter $c^*_\gamma(t)$, shown by the solid line. Secondly, we observe that the graphs, except for $\gamma = 1$, intersect in approximately the same point t^*, which implies that, in this point, the optimal consumption rate is independent on the risk preferences, given by γ. The curve for $\gamma = 1$ differs from the other curves due to the fact that at $\gamma = 1$ the utility function is logarithmic (see Eq. (8.36)).

These two observations together indicate that the value of $c^*_\gamma(t)$, calculated for $\gamma \to \infty$, may crosse $c^*_\gamma(t)$ calculated for fixed γ, in t^*. We check this empirical assumption by determining

$$\begin{aligned}\lim_{\gamma\to\infty} c^*_\gamma(t) &= \lim_{\gamma\to\infty} c^*_\gamma(0) e^{\frac{r-\alpha}{\gamma}t} \left(p_t\right)^{1/\gamma} \\ &= \lim_{\gamma\to\infty} c^*_\gamma(0) = \frac{\pi_0/r(e^{rt}-1)+We^{rt}}{e^{rt}\lim_{\gamma\to\infty} I_\gamma},\end{aligned} \tag{8.43}$$

where

$$\begin{aligned}\lim_{\gamma\to\infty} I_\gamma &= \lim_{\gamma\to\infty} \int_0^t \left(\underline{\tau}_x e^{\Lambda s}\mathbf{1}\right)^{1/\gamma} e^{\frac{r(1-\gamma)-\alpha}{\gamma}s} ds \\ &= \int_0^t e^{-rs} ds = (1-e^{-rt})\frac{1}{r}.\end{aligned} \tag{8.44}$$

Eq. (8.43) and Eq. (8.44) together result in

$$\lim_{\gamma\to\infty} c^*_\gamma(t) = \pi_0 + Wr/(1-e^{-rt}).$$

Therefore, in order to determine t^*, we solve numerically

$$|c^*_{\hat{\gamma}}(t) - \lim_{\gamma\to\infty} c^*_\gamma(t)| \le \varepsilon,$$

for a fixed $\hat{\gamma}$ and ε. For example, we find that if $\hat{\gamma} = 3$ and $\varepsilon = 10^{-2}$, then t^* is about 13, which is confirmed by Fig. 8.4.

In Fig. 8.5 we illustrate $c^*(t)$ for different initial ages x and $\gamma = 19$. One sees from the figure, that the younger the individual at the beginning, the more uniform the consumption rate becomes. We again observe that the curves intersect at about the same time.

There are several research directions to investigate here. One direction is to incorporate a stochastic interest rate, this would allow us to examine the impact of the economy on the individual consumption. Another direction is to incorporate stochasticity in the future mortality rates. This would help us to estimate the effect of longevity problem on individual households. There are many studies (see, for example, [33]) that allow to solve the optimal consumption problem (8.34) if the mortality rates depend on a Brownian motion. To keep the connection with the PH-aging model one may implement the *stochastic aging model*, introduced in Section 2.5.

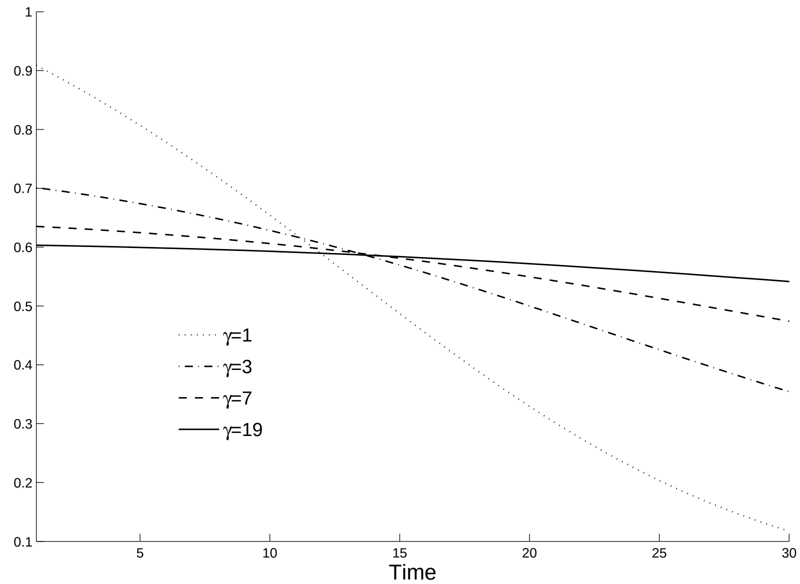

Figure 8.4: Optimal rate of consumption: effect of γ
Parameters: $x = 65$, $\alpha = 0.05$, $W = 10, r = 0.05$, $\pi_0 = 0$, SW1911M

Health economics

In the book by Drummond et al. [20] it is explained why cost-effectiveness and cost-utility analyses of health care are important issues in health economics. The analyses help to solve such problems as decision making regarding the production of new drugs, or determining the optimal treatment strategies. The efficiency or utility criteria often depend on the estimated quality and quantity of years lived after taking a treatment. Clearly, this implies that it is important to take into account the health state development of an individual before and after the treatment. For this purpose, it is quite common (see, for example, Castelli et al. [17]) to use a Markovian model. However, as we have mentioned in the introduction to Part III, such models can be improved with the help of the PH-aging model. Below we give some indications for future research on this topic.

Consider the cost-effectiveness problem of the production and of the consumption of health care. An example of the widely used criterion of effectiveness is the *quality-adjusted life year (QALY)* (see Drummond et al. [20]). QALY is the measure of disease burden, which takes into account both the quality of life and the number of

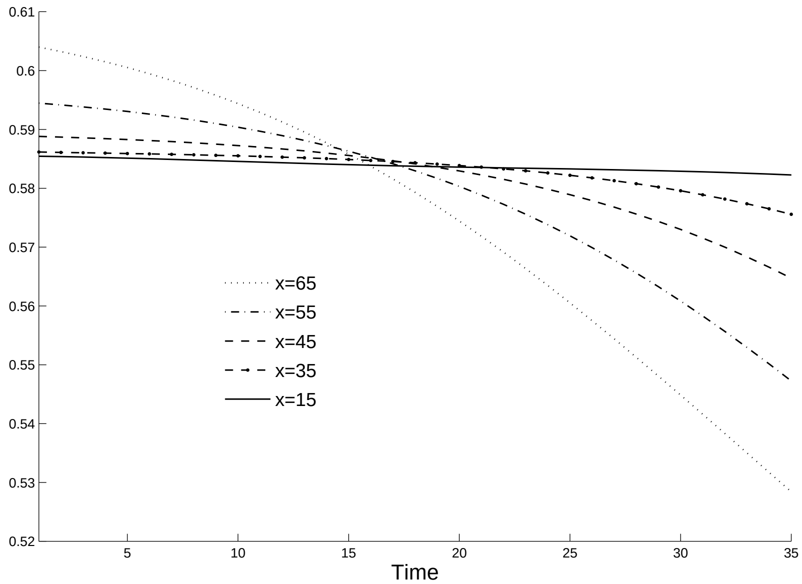

Figure 8.5: Optimal rate of consumption: *impact of health*
Parameters: $\gamma = 19$, $\alpha = 0.05$, $W = 10, r = 0.05$, $\pi_0 = 0$, SW1911M

years lived. In order to compute its value, one associates a utility value to the states of health.

Assume that we need to choose the optimal treatment strategy for an individual suffering from a certain disease. Furthermore, assume that the individual/government has a limited budget, and therefore can not automatically choose the most effective strategy. We suggest the following steps to determine QALY for possible treatment strategies.

The problem of determining QALY for a treatment strategy (k) may be reduced to the computation of the distribution of

$$J^{(k)} = \int_0^L f(t, \phi_t) dt,$$

where L is the remaining lifetime after the treatment, $f(t, \phi_t)$ is a function of time t and the health state ϕ_t at time t. One may also compute the distribution of the net present value of costs for treatment (k), given by

$$C^{(k)} = \int_0^L v^k r_{\phi_t} dt,$$

where v is a discount factor and r_{ϕ_t} is the cost rate for the state ϕ_t. Thus, in order to choose an efficient strategy one may compute the distributions of $J^{(k)}$ and $C^{(k)}$ for all k and compare all outcomes.

Technically, to compute $J^{(k)}$ and $C^{(k)}$ one should determine L, $f(t, \phi_t)$ and r_{ϕ_t}. We suggest to describe L by a phase-type distribution, similarly as it is done in the PH-aging model of Lin and Liu [40]. We believe that one needs to adapt the PH-aging model to describe the health state progression of an individual subject to the disease and the treatment (k). It is important to collaborate with experts in medicine to learn about the progression of the disease: this would help to determine the structure of the new transition matrix and the function $f(t, \phi_t)$. The cost rates r_{ϕ_t} one may obtain using the procedure that we have developed in Section 2.3.

Due to the assumption that L is of phase-type, in order to obtain the distributions of $J^{(k)}$ and $C^{(k)}$ one may use the approaches that we have developed in Part III for the net present value of health care costs.

Appendix A

Useful algebra

A.1 Kronecker operations

The *Kronecker product* of two matrices A and B, denoted by $A \otimes B$, is an operation resulting in a block matrix. If A is a $m \times n$ matrix and B is a $p \times q$ matrix, then $A \otimes B$ is a $mp \times nq$ block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & ... & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & ... & a_{mn}B \end{bmatrix}. \tag{A.1}$$

The Kronecker product is bilinear and associative:

$$\begin{aligned} &A \otimes (B + C) = A \otimes B + A \otimes C, \\ &(A + B) \otimes C = A \otimes C + B \otimes C, \\ &(kA) \otimes B = A \otimes (kB) = k(A \otimes B), \\ &(A \otimes B) \otimes C = A \otimes (B \otimes C), \end{aligned} \tag{A.2}$$

where A, B and C are matrices and k is a scalar.

The *Kronecker sum* of two matrices A and B of size $n \times n$ and $m \times m$, respectively, is denoted by $A \oplus B$ and is defined as

$$A \oplus B = A \otimes I_m + I_n \otimes B, \tag{A.3}$$

where I_k is the $k \times k$ identity matrix.

The *Kronecker exponentiation* is given by

$$e^{A \oplus B} = e^A \otimes e^B, \tag{A.4}$$

where e^X is the matrix exponential of the matrix X.

Using the indicated properties one can prove that

$$\begin{aligned} \underline{\alpha} e^A \mathbf{1} \underline{\beta} e^B \mathbf{1} &= (\underline{\alpha} \otimes \underline{\beta})(e^A \otimes e^B)(\mathbf{1} \otimes \mathbf{1}) \\ &= (\underline{\alpha} \otimes \underline{\beta}) e^{A \oplus B}(\mathbf{1} \otimes \mathbf{1}). \end{aligned} \tag{A.5}$$

A.2 Some distributions

A random variable X that is *gamma-distributed* with shape k and scale θ is denoted by $X \sim \Gamma(k, \theta)$ and has the following properties:

- $E[X] = k\theta$;
- $Var[X] = k\theta^2$;
- the probability density function is given by

$$f(x; k, \theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}, \quad x > 0, \quad k, \theta > 0. \tag{A.6}$$

Here,

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt.$$

A *lognormal random variable* is a random variable whose logarithm is normally distributed. A lognormal random variable X with parameters μ and σ can be represented as $X = e^{\mu + \sigma Y}$, where Y is a standard normal random variable, and has the following properties:

- $E[X] = e^{\mu + \frac{1}{2}\sigma^2}$;
- $Var[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$;
- the probability density function is given by

$$\phi_l(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\dfrac{(\ln x - \mu)^2}{2\sigma^2}}; \tag{A.7}$$

- the cumulative distribution functions is

$$\Phi_l(x; \mu, \sigma^2,) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right), \tag{A.8}$$

where $\Phi(x)$ is a cumulative distribution function of a standard normal random variable.

A.3 Stochastic processes

Let $X = (X(t), t \geq 0)$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P)$. We refer to Applebaum [2] to introduce a *Lévy process* X as follows:

i. $X(0) = 0$ almost surely;

ii. X has independent and stationary increments: for each $n \in \mathbb{N}$ and each $0 \leq t_0 < t_1 < ... < t_{n+1} < \infty$ the random variables $(X(t_{j+1}) - X(t_j), 0 \leq j \leq n)$ are independent, and the random variables $X(t_{j+1}) - X(t_j)$ and $X(t_{j+1} - t_j) - X(0)$ are equal in distribution;

iii. X is stochastically continuous: for all $a > 0$ and for all $s \geq 0$

$$\lim_{t \to s} \mathrm{P}[\,|X(t) - X(s)| > a\,] = 0.$$

A *standard Brownian motion* in $\mathbb{R}$ is a Lévy process $(W(t), t \geq 0)$ such that

- $W(t) \sim \mathbb{N}(0, t)$ for each $t \geq 0$;
- $W(t)$ has continuous sample paths.

A *Brownian motion with drift* is a Lévy process $(B(t), t \geq 0)$ such that

$$B(t) = \mu t + \sigma W(t), \tag{A.9}$$

where $\mu, \mu \in \mathbb{R}$ is called a *drift* and $\sigma, \sigma \in \mathbb{R}$ is a *diffusion coefficient*. It is easy to see that $B(t) \sim \mathbb{N}(\mu t, \sigma^2 t)$.

A *geometric Brownian motion* can be defined as a continuous time stochastic process $(S(t), t \geq 0)$ of the form

$$S(t) = S(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)}. \tag{A.10}$$

Thus, $S(t)$ is a lognormal random variable with mean $S(0)e^{\mu t}$ and variance

$$S^2(0)e^{2\mu t}(e^{\sigma^2 t} - 1).$$

A *subordinator* is a one-dimensional Lévy process that is non-decreasing (almost surely). One example of the subordinator is a *gamma process*. A subordinator $\gamma_t, t \geq 0$ is called a gamma process with parameters $k, \theta > 0$ if γ_t is a gamma distributed random variable with probability density function $f(x; kt, \theta)$, given by Eq. (A.6).

A.4 Convolution

Suppose that X and Y are independent random variables with distribution functions F and G, respectively. Then the distribution of $X + Y$, denoted by $F * G$ and called the *convolution* of F and G, is given by

$$(F * G)(a) = \int_{-\infty}^{\infty} F(a - y)dG(y). \tag{A.11}$$

If X and Y are discrete random variables with probability mass functions f and g, respectively. Then the distribution of $X + Y$, denoted by $f * g$ and called the *convolution* of f and g, is given by

$$(f * g)(a) = \sum_{k=-\infty}^{\infty} f(k)g(a - k). \tag{A.12}$$

A.5 Panjer recursion

The probability distribution of a counting random variable N belongs to the *Panjer class*, if

$$\mathrm{P}[\, N = k \,] = p_k = (a + b/k)p_{k-1}, \quad k \geq 1,$$

for some a and b, which fulfill $a + b \geq 0$. The initial value p_0, is determined such that

$$\sum_{k=0}^{\infty} p_k = 1.$$

The *Panjer recursion* is an algorithm to compute the probability distribution of a compound random variable

$$S = \sum_{i=1}^{N} X_i,$$

where X_i are i.i.d. random variables and independent of N. S and X_i have a distribution on a lattice $h\mathbb{N}_0$ of width $h > 0$

$$f_k = \mathrm{P}[\, X_i = hk \,], \quad g_k = \mathrm{P}[\, S = hk \,].$$

The algorithm gives a recursion to compute g_k:

- if $a = 0$,
$$\begin{cases} g_0 = p_0 e^{f_0 b}, \\ g_k = \displaystyle\sum_{j=1}^{k} \frac{bj}{k} f_j g_{k-j}; \end{cases}$$

- if $a \neq 0$,
$$\begin{cases} g_0 = \dfrac{p_0}{(1 - f_0 a)^{1+b/a}}, \\ g_k = \dfrac{1}{1 - f_0 a} \displaystyle\sum_{j=1}^{k} (a + \frac{bj}{k}) f_j g_{k-j}. \end{cases}$$

If N is a Poisson random variable with parameter λ, then

$$a = 0, \quad b = \lambda, \quad p_0 = e^{-\lambda},$$

and the Panjer recursion takes form

$$\begin{cases} g_0 = e^{-\lambda(1-f_0)} \\ g_k = \displaystyle\sum_{j=1}^{k} \frac{\lambda j}{k} f_j g_{k-j}. \end{cases} \tag{A.13}$$

Bibliography

[1] D. Acemoglu, G. Egorov, and K. Sonin. Political model of social evolution. *Proceedings of the National Academy of Sciences*, 108:21292–21296, 2011.

[2] D. Applebaum. *Lévy Processes and Stochastic Calculus.* Cambridge university press, 2009.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[4] S. Asmussen and H. Albrecher. *Ruin probabilities.* World Scientific Publishing Co., 2010.

[5] S. Asmussen, O. Nerman, and M. Olsson. Fitting Phase-Type distributions via the EM Algorithm. *Scandinavian Journal of Statistics*, 23(4):419–441, 1996.

[6] A. Badescu, L. Breuer, A. D. S. Soares, G. Latouche, M.-A. Remiche, and D. Stanford. Risk Processes Analyzed as Fluid Queues. *Scandinavian actuarial journal*, 00:1–16, 2004.

[7] A. L. Badescu and D. Landriault. Recursive calculation of the dividend moments in a multi-threshold risk model. *North American Actuarial Journal*, 12(1), 2008.

[8] P. Barrieu, H. Bensusan, N. E. Karoui, C. Hillairet, S. Loisel, C. Ravanelli, and Y. Salhi. Understanding, Modeling and Managing Longevity Risk: Key Issues and Main Challenges. *Scandinavian actuarial journal*, 3:203–231, 2012.

[9] L. Bertschi, S. Ebeling, and A. Reichlin. Dynamic Asset Liability Management: a profit testing model for Swiss pension funds. *ASTIN Colloquium International Actuarial Association*, 2003.

[10] N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones, and C. J. Nesbitt. *Actuarial Mathematics*. The Society of Actuaries, 1997.

[11] L. Breuer. Occupation times for Markov modulated Brownian motion. *Journal of Applied Probability*, 49(2):549–565, 2012.

[12] D. R. Brillinger. The natural variability of vital rates and associated statistics. *Biometrics*, 42:693–734, 1986.

[13] P. L. Brockett. Information theoretic approach to actuarial science: a unification and extension of relevant theory and applications. *Transactions of Society of Actuaries*, 43:73–135, 1991.

[14] B. Bukiet and E. R. Harold. A Markov chain approach to baseball. *Operations research*, 45(1):14–23, 1997.

[15] A. J. G. Cairns, D. Blake, and K. Dowd. A two factor model for stochastic mortality with parameter uncertainty: theory and calibration. *The journal of risk and insurance*, 73(4):687–718, 2006.

[16] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41(1):29–59, 2011.

[17] C. Castelli, C. Comberscure, Y. Foucher, and J.-P. Daures. Cost-effectiveness analysis in colorectal cancer using a semi-markov model. *Statistics in medicine*, 26:5557–5571, 2007.

[18] E. Çinlar. *Introduction to Stochastic Processes*. Prentice–Hall, Englewood Cliffs, NJ, 1975.

[19] G. E. Cho and C. D. Meyer. Markov chain sensitivity measured by mean first passage times. *Linear Algebra and its Applications*, 316(1-3):21–28, 2000.

[20] M. F. Drummond, M. J. Sculpher, G. W. Torrance, B. J. O'Brien, and G. L. Stoddart. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford university press, 2005.

[21] D. Duffie and J. Pan. An Overview of Value at Risk. *The Journal of Derivatives*, 4(3):7–49, 1997.

[22] K.-T. Eisele. Recursions for compound phase distributions. *Insurance: Mathematics and Economics*, 38:149–156, 2006.

[23] J. Elliot and A. Swishchuk. Pricing options and variance swaps in Markov modulated Brownian markets. *Hidden Markov Models in Finance*, 104:45–68, 2007.

[24] D. M. Franz. Markov chains as tools for jazz improvisation analysis. Master's thesis, Virginia Polytechnic Institute and State University, 1998.

[25] J. C. Gardiner, Z. Luo, C. J. Bradley, C. M. Sirbu, and C. W. Given. A dynamic model for estimating changes in health status and costs. *Statistics in medicine*, 25:3648–3667, 2006.

[26] H. U. Gerber. *Life insurance mathematics*. Springer-Verlag Berlin Heidelberg and Swiss Association of Actuaries Zurich, 1990.

[27] S. Gerhold, U. Schmock, and R. Warnung. A generalization of panjer's recursion and numerically stable risk aggregation. *Finance and Stochastics*, 14:81–128, 2010.

[28] M. Govorun and G. Latouche. Modeling the effect of health: phase-type approach. *Submitted for publication*, 2013.

[29] M. Govorun, G. Latouche, and M.-A. Remiche. Profit test model for pension funds using matrix-analytic modeling. *Proceedings of the Actuarial and Financial Mathematics Conference — Interplay Between Finance and Insurance*, pages 87–94, 2011.

[30] M. Govorun, G. Latouche, and M.-A. Remiche. Profits and risks of pension plans. *Performance Evaluation Review*, 39:41, 2012.

[31] J. M. Hoem. Markov chain models in life insurance. *Blätter der Deutschen Gesellschaft für Versicherungsmathematik*, IX:91–107, 1969.

[32] A. Horvath and M. Telek. PhFit: A general phase-type fitting tool. *Proc. of Performance TOOLS 2002*, 2002.

[33] H. Huang, M. A. Milevsky, and T. S. Salisbury. Optimal retirement consumption with a stochastic force of mortality. *Insurance: Mathematics and Economics*, 51:282–291, 2012.

[34] J. Janssen and R. Manca. Semi-Markov Modelisation for the Financial Management of Pension Funds. *4th AFIR International Colloquium*, pages 1369 – 1387, 1994.

[35] B. L. Jones. Actuarial calculations using a Markov model. *Transactions of Society of Actuaries*, 46, 1994.

[36] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* ASA-SIAM, 1999.

[37] R. D. Lee and L. R. Carter. Modelling and forecasting U.S. Mortality. *Journal of the American Statistical Association*, (87):659–675, 1992.

[38] M. Leppisaari. Managing longevity risk with longevity bonds. Technical report, Department of Mathematics and Systems Analysis, Faculty of Information and Natural Sciences, Helsinki University of Technology, August 2008.

[39] X. S. Lin. *Introductory Stochastic Analysis for Finance and Insurance.* John Wiley & Sons, Inc., 2006.

[40] X. S. Lin and X. Liu. Markov aging process and Phase-Type Law of mortality. *North American Actuarial Journal*, 11:92–109, 2007.

[41] X. S. Lin and X. Liu. A subordinated markov model for stochastic mortality. *European Actuarial Journal*, 2(1):105–127, 2012.

[42] X. Liu. *Stochastic Mortality Modelling.* PhD thesis, Department of Statistics, University of Toronto, 2008.

[43] S. Loisel and X. Milhaud. From deterministic to stochastic surrender risk models: impact of correlation crises on economic capital. *European Journal of Operational Research*, 214(2):348–357, 2011.

[44] U. Mettler. Projecting pension fund cash flows. Technical Report 1, National Centre of Competence in Research Financial Valuation and Risk Management, Zurich, Switzerland, 2005.

[45] M. F. Neuts. *Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach*. Johns Hopkins University Press, Baltimore, 1981.

[46] R. Norberg. The Markov Chain Market. *Astin Bulletin*, 33(2):265–287, 2003.

[47] J. Oeppen and J. W. Vaupel. Broken limits to life expectancy. *SCIENCE*, 296, 2002.

[48] M. Olsson. The EMpht-programme. Technical report, Department of mathematics, Chalmers University of Technology, and Gothenburg University, June 1998.

[49] H. H. Panjer. Recursive evaluation of a family of compound distributions. *Astin Bulletin*, 12:22–26, 1981.

[50] E. Pitacco. Survival models in actuarial mathematics: from Halley to longevity risk. *Invited talk at 7th International Congress Insurance Mathematics and Economics*, 2003.

[51] A. H. Pollard. Random mortality fluctuations and the binomial hypothesis. *Journal of the Institute of Actuaries*, 96:251–264, 1970.

[52] S. M. Ross. *Stochastic Processes*. John Wiley, New York, 1996.

[53] M. Schröder. On the integral of geometric Brownian motion. *Advances in Applied Probability*, 35(1):159–183, 2003.

[54] W. F. Scott. *Life Insurance Mathematics*. Department of Mathematical Sciences, King's College, University of Aberdeen, Aberdeen, U.K., 1999.

[55] E. Seneta. Sensitivity of finite Markov chains under perturbation. *Statistics&Probability Letters*, 17:163–168, 1993.

[56] I. C. Smart. Pricing and Profitability in a Life Office. *Journal of the Institute of Actuaries*, 104, 1977.

[57] D. Stanford, F. Avram, A. Badescu, L. Breuer, A. da Silva Soares, and G. Latouche. Phase-type approximations to finite-time ruin probabilities in the Sparre-Andersen and stationary renewal risk models. *Methodology and computing in applied probability*, 10:381–408, 2008.

[58] B. Sundt and W. S. Jewell. Further results on recursive evaluation of compound distributions. *Astin Bulletin*, 12:27–39, 1981.

[59] J. Thomsen and J. V. Andersen. Longevity Bonds – a financial market instrument to manage longevity risk. *Denmark: Monetary Review - 4th Quarter 2007.*

[60] P. P. Wakker. Explaining the characteristics of the power (crra) utility family. *Health Economics*, 17:1329–1344, 2008.

[61] Website. Canadian Institute for Health Information http://www.cihi.ca.

[62] Website. The Human Mortality Database http://www.mortality.org.

[63] Website. Wikipedia http://en.wikipedia.org.

[64] H. Wolthuis. *Life Insurance Mathematics (The Markovian Model).* Caire Education Series 2, 1994.

[65] M. Yaari. Uncertain lifetime, life insurance and the theory of the consumer. *The Review of Economic Studies*, 32(2):137–150, 1965.

[66] M. Yor. On some exponential functionals of Brownian motion. *Advances in Applied Probability*, 24(3):509–531, 1992.

[67] X. Zhao and X. Zhou. Estimation of medical costs by copula models with dynamic change of health status. *Insurance: Mathematics and Economics*, 2012.

Index

Pension and Health Insurance: Phase-Type Modeling

Maria Govorun

Depuis longtemps les modèles de type phase sont utilisés dans plusieurs domaines scientifiques pour décrire des systèmes qui peuvent être caractérisés par différents états. Les modèles sont bien connus en théorie des files d'attentes, en économie et en assurance.

La thèse est focalisée sur différentes applications des modèles de type phase en assurance. En particulier, le modèle de Lin et Liu en 2007 est intéressant, parce qu'il décrit le processus de vieillissement de l'organisme humain. La durée de vie d'un individu suit une loi de type phase et les états de ce modèle représentent des états de santé. Le fait que le modèle prévoit la connexion entre les états de santé et l'âge de l'individu le rend très utile en assurance. Les résultats principaux de la thèse sont des nouveaux modèles et méthodes en assurance pension et en assurance santé qui utilisent l'hypothèse de la loi de type phase pour décrire la durée de vie d'un individu.

En assurance pension le but est d'estimer la profitabilité d'un fonds de pension. Pour cette raison, on construit un modèle *profit-test* qui demande la modélisation de plusieurs caractéristiques. On décrit l'évolution des participants du fonds en adaptant le modèle du vieillissement aux causes multiples de sortie. L'estimation des profits futurs exige qu'on détermine les valeurs des cotisations pour chaque état de santé, ainsi que l'ancienneté et l'état de santé initial pour chaque participant. Cela nous permet d'obtenir la distribution de profits futurs et de développer des méthodes pour estimer les risques de longevité et de changements de marché. De plus, on suppose que la diminution des taux de mortalité pour les pensionnés influence les profits futurs plus que pour les participants actifs. C'est pourquoi, pour évaluer l'impact de changement de santé sur la profitabilité, on modélise séparément les profits venant des pensionnés.

En assurance santé, on utilise le modèle de type phase pour calculer la distribution de la valeur actualisée des coûts futurs de santé. On développe des algorithmes récursifs qui permettent d'évaluer la distribution au cours d'une période courte, en utilisant des modèles fluides en temps continu, et pendant toute la durée de vie de l'individu, en construisant des modèles en temps discret. Les modèles en temps discret correspondent à des hypothèses différentes qu'on fait pour les coûts: dans un modèle on suppose que les coûts de santé sont indépendants et identiquement distribués et ne dépendent pas du vieillissement de l'individu; dans les autres modèles on suppose que les coûts dépendent de son état de santé.