Multi-state modeling in the mathematics of life insurance: meditations and applications

Christian Furrer

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

PFA Pension

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen.

Industrial PhD Thesis by:

Christian Furrer
Vognmandsmarken 29, 2. th.
DK-2100 Copenhagen Ø
christian@furrer.dk
http://christian.furrer.dk

Assessment committee:

Professor Thomas Mikosch (chairperson)
University of Copenhagen

Professor Boualem Djehiche
Royal Institute of Technology, Sweden

PhD Matthias Albrecht Fahrenwaldt
Federal Financial Supervisory Authority, Germany

Supervisors:

Professor Mogens Steffensen
University of Copenhagen

PhD Kristian Buchardt
PFA Pension

Professor Niels Richard Hansen
University of Copenhagen

Adjunct Professor Thomas Møller
AP Pension & University of Copenhagen

PhD Peter Holm Nielsen
PFA Pension

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen on 27 August 2020. It has been written as part of an Industrial PhD project with File No. 7038-00007B funded jointly by PFA Pension and Innovation Fund Denmark.

Thomas Møller supervised until he changed affiliation from PFA Pension to AP Pension in August 2018, while Peter Holm Nielsen supervised from September 2018.

ISBN: 978-87-7125-036-7.

Abstract

This thesis consists of a series of independent investigations pertaining primarily to multi-state modeling in the mathematics of life insurance. First, we study the dynamics of state-wise prospective reserves in the presence of non-monotone information. The corresponding main result consists of a generalization of the stochastic Thiele equations. Next, we present and discuss a series of questions concerning the representation and computation of expected accumulated cash flows in the presence of bonus, incidental policyholder behavior, and double stochasticity, respectively. Central contributions include the derivation of procedures for the computation of the market value of bonus payments, establishment of links between measure changes and scaling factors resulting from the exercise of policyholder options, and the comparison of pros and cons of various concepts of forward transition rates. Following this, we study experience rating for multi-state life insurance by applying empirical Bayes methods to a multivariate frailty extension of the classic setup. The thesis concludes with an extension of the quadratic hedging approach known as risk-minimization to allow for taxes and expenses.

Preface

> *"Sigma: But then nothing is settled. We can't stop now.*
> *Teacher: I sympathise. This latest stage will have important feedbacks to our discussion. But a scientific inquiry 'begins and ends with problems'. [Leaves the classroom.]*
> *Beta: But I had no problems at the beginning! And now I have nothing but problems!"*
>
> — Imre Lakatos, *Proofs and Refutations*

This thesis has been prepared in partial fulfillment of the requirements for the PhD degree at the Department of Mathematical Sciences, Faculty of Science, University of Copenhagen. The work has been carried out between September 2017 and August 2020 as an Industrial PhD project within Innovation Fund Denmark's program and with PFA Pension as the industrial partner. I was supervised by Professor Mogens Steffensen (University of Copenhagen), PhD Kristian Buchardt (PFA Pension), Professor Niels Richard Hansen (University of Copenhagen), Adjunct Professor Thomas Møller (PFA Pension until August 2018, then AP Pension; University of Copenhagen throughout), and PhD Peter Holm Nielsen (PFA Pension). Thomas Møller supervised until August 2018, while Peter Holm Nielsen supervised from September 2018.

The thesis consists of manuscripts that have been produced as part of my studies. The manuscripts constitute stand-alone scientific contributions and should be read as such. This very nature of the thesis leads to significant discrepancies in especially notation across chapters. An overview and contextualization of the main contributions of the thesis and their interconnections can be found in Chapter 1. Minor differences between the contents of a chapter and the corresponding manuscript might exist. I take full responsibility for any typographical or mathematical errors.

Ever since defending my master's thesis in actuarial mathematics at the University of Copenhagen three years ago, I have been blessed with abundant opportunities for personal and professional growth. I have not been afraid to speak my mind, and I have tried to listen. I have overcome some challenges, and some have overcome me. According to the late Leonard Cohen: *"That's how the light gets in"*. Whatever it might hold, the future excites me.

Acknowledgments

I should like to thank Thomas Møller for encouraging me to undertake the studies and Innovation Fund Denmark and PFA Pension for funding the project. Thomas, your mentoring leading up to and during the early stages of my PhD studies has been invaluable.

I thank Mogens Steffensen and Kristian Buchardt for three years of thoughtful and supportive supervision, and I thank Niels Richard Hansen for instructive discussions and Peter Holm Nielsen for continuously supporting the project. Kristian, your commitment to the project and your interest in my personal and professional development have been astonishing, and you have gone out of your way to help me succeed. Learning from you and with you has been wonderful, and I consider myself fortunate to have been given this opportunity. I hope the future holds opportunities for collaboration in regards to both research and teaching. Mogens, your supervision concerning project and time management in particular has been vital for the success of the project. More than most you stand by your word, and you have had my back when it really mattered. I enjoyed our walk and talks in Fælledparken and wish for them to be the first of many yet to come.

In the late summer and early fall of 2018 and again in the summer of 2019, I had the pleasure of visiting Marcus C. Christiansen at the Carl von Ossietzky University of Oldenburg. I should like to thank Marcus, Marius, Peter Ruckdeschel, Angelika, and Peter Krug for a warm welcome and a wonderful time in Oldenburg. Marcus, I am grateful for having been given the opportunity to work with you and learn from you. Your work ethic and willingness to share ideas and insights are incredible and a great inspiration to me. I look forward to continuing our collaboration.

I should like to express my gratitude to Mario Wüthrich for inviting me to give a talk in the seminar on Insurance Mathematics and Stochastic Finance at ETH Zürich in December 2019. I should also like to thank The German Actuarial Society and The German Society for Insurance and Financial Mathematics for awarding me the prestigious Gauss Prize for my first paper which appeared in the European Actuarial Journal in 2019.

Many thanks to my co-authors Jamaal Ahmad, Kristian Buchardt, Marcus C. Christiansen, Thomas Møller, and Mogens Steffensen for fruitful collaborations. Thanks to my colleagues at PFA Pension, especially Lars, for your support and encouragements. I am going to miss you. Thanks to my colleagues at the Department of Mathematical Sciences, especially Jesper, for your open doors and open minds. I look forward to strengthening our collaboration.

I thank Jessie for proofreading, of course, but foremost for your compassion and kindness. Thanks to Brix for always listening, and thanks to Ann Sofie for sometimes doing quite the opposite.

Last but not least, to Frantz, Laura, Far, and Mor: nothing but my deepest gratitude and affection. You have taught me to toil relentlessly and to love unconditionally. *"Zum Augenblicke dürft' ich sagen: Verweile doch, du bist so schön!"*. Nonetheless, I shall promise to try to lose my footing momentarily.

Christian Furrer
Copenhagen, August 2020

This final version of my thesis differs only from the previous version submitted to the PhD School on 27 August 2020 in that an ISBN has been provided and a few typographical errors have been corrected.

Christian Furrer
Copenhagen, October 2020

List of papers

Besides Chapter 1 and Chapter 4, which have been prepared specifically for this thesis, each of the five remaining chapters presents a self-contained and submitted or already published manuscript according to the following scheme:

Chapter 2: Christiansen, M.C. and C. Furrer (2020). *Dynamics of state-wise prospective reserves in the presence of non-monotone information.* arXiv: 2003.02173.

Chapter 3: Ahmad, J., K. Buchardt, and C. Furrer (2020). *Computation of bonus in multi-state life insurance.* arXiv: 2007.04051.

Chapter 5: Buchardt, K., C. Furrer, and M. Steffensen (2019). Forward transition rates. *Finance and Stochastics* **23**(4), pp. 975–999. doi: 10.1007/s00780-019-00397-0.

Chapter 6: Furrer, C. (2019). Experience rating in the classic Markov chain life insurance setting. *European Actuarial Journal* **9**, pp. 31–58. doi: 10.1007/s13385-019-00190-5.

Chapter 7: Buchardt, K., C. Furrer, and T. Møller (2020). Tax- and expense-modified risk-minimization for insurance payment processes. To appear in *Scandinavian Actuarial Journal.* doi: 10.1080/03461238.2020.1790413.

There might be minor differences between the contents of a chapter and the corresponding manuscript.

Summary

This thesis consists of several independent investigations pertaining primarily to multi-state modeling in the mathematics of life insurance. Following Chapter 1, which sets the stage and provides an overview of the main contributions of the thesis and their interrelations, the investigations are presented in Chapters 2–7, with each chapter forming a single investigation.

In broad terms, the thesis is concerned with conceptual and computational challenges arising in multi-state life insurance. We apply and extend methods from probability theory, specifically the theory of stochastic processes, and mathematical finance to solve actuarial problems of theoretical and practical importance. The focus is first and foremost on abstract concepts, but multiple examples and case studies also illustrate the applicability of our methods and results in actuarial practice.

In Chapter 2, which contains the manuscript *Christiansen and Furrer (2020)*, we discuss valuation in the presence of non-monotone information. Non-monotonicity arises if the insurer does not have access to or does not desire to utilize all possible information concerning the states of the insured, e.g. due to legal constraints resulting from privacy law. By adopting an infinitesimal approach, we derive stochastic differential equations describing the dynamics of state-wise prospective reserves. To this end, we clarify definitions and properties of different notions of state-wise prospective reserves. A case study involving information discarding upon and after stochastic retirement exemplifies the methods and results.

Chapter 3–5 are concerned with the representation and computation of expected accumulated cash flows in the presence of bonus, incidental policyholder behavior, and double stochasticity, respectively. In Chapter 3, which contains the manuscript *Ahmad, Buchardt, and Furrer (2020)*, we consider with-profit contracts and the bonus scheme *additional benefits*, where dividends are used to buy extra benefits. Requiring the dividend strategy to be affine in the number of additional benefits, we derive a procedure for the computation of the market value of bonus payments which efficiently combines simulation of financial risk with classic methods for insurance risk. Special attention is given to the case where the number of additional benefits only depends on financial risk – building a bridge between collective and individual points

of view. Incidental policyholder behavior leads to payments scaled according to the exercise times of policyholder options. In Chapter 4, we establish a link between scaling factors and measure changes using supermartingales as Radon-Nikodym derivatives. This link can be used to conveniently derive backward and forward methods for the computation of prospective reserves. Chapter 5, which contains the paper *Buchardt, Furrer, and Steffensen (2019)*, studies forward transition rates in doubly stochastic Markov chain models. We establish a theoretical framework, propose a new concept, and compare it to earlier proposals in the literature: marginal and state-wise forward transition rates.

Contrary to the remaining investigations, Chapter 6, which contains the paper *Furrer (2019)*, focuses on statistical rather than probabilistic aspects of multi-state modeling, namely experience rating for multi-state life insurance. To this end, we apply empirical Bayes methods to multivariate frailty extensions of classic Markov chain models. Special attention is given to the case where the group effects are mutually independent and Gamma-distributed, where the classic link to Poission regressions is replaced by a link to multivariate negative binomial regressions. The methods and results are illustrated by a numerical example for disability insurance using simulated data.

Taxes on investment returns lead to insurance payments which depend on the investment strategy. Consequently, classic methods for market-consistent valuation do not apply. In Chapter 7, which contains the paper *Buchardt, Furrer, and Møller (2020)*, we consider quadratic hedging of insurance payment processes in the presence of taxes and expenses. The chapter differs from the remaining investigations by not having multi-state modeling as its focal point. We propose the criterion of tax- and expense-modified risk-minimization, which takes into account the effect of taxes and expenses on the time value of money. We establish existence and uniqueness of an optimal investment strategy related to the Galtchouk-Kunita-Watanabe decomposition of the intrinsic value process associated with a tax- and expense-modified payment process. The investigation concludes with an application of tax- and expense-modified risk-minimization aimed at multi-state life insurance.

Resumé

Denne afhandling består af flere uafhængige undersøgelser, som har at gøre med flertilstandsmodellering i livsforsikringsmatematik. I forlængelse af Kapitel 1, der sætter scenen og giver et overblik over afhandlingens hovedbidrag og deres indbyrdes sammenhænge, præsenteres undersøgelserne i Kapitel 2–7, idet hvert kapitel udgør en enkeltstående undersøgelse.

Overordnet set omhandler denne afhandling konceptuelle og beregningsmæssige udfordringer i flertilstandslivsforsikring. Vi anvender og udvider metoder fra sandsynlighedsteori, specifikt teorien om stokastiske processer, og finansmatematik for at løse aktuarmæssige problemer af teoretisk og praktisk betydning. Fokus er først og fremmest abstrakte koncepter, men adskillige eksempler og casestudier er også med til at illustrere anvendeligheden af vores metoder og resultater i aktuarmæssig praksis.

I Kapitel 2, som indeholder manuskriptet *Christiansen og Furrer (2020)*, diskuterer vi værdiansættelse, når den tilgængelige information er ikke-monoton. Ikke-monotonicitet opstår, hvis forsikringsselskabet ikke har adgang til eller ikke ønsker at udnytte al information om de forsikredes tilstande, fx på grund af juridiske begrænsninger som følge af privatlivsret. Ved at anlægge en infinitesimal fremgangsmåde udleder vi stokastiske differentialligninger, som beskriver dynamikken af tilstandsvise prospektive reserver. Til dette formål afklarer vi definitioner og egenskaber af forskellige opfattelser af tilstandsvise prospektive reserver. Et casestudie, som involverer sletning af information ved og efter stokastisk pensionering, eksemplificerer metoderne og resultaterne.

Kapitel 3–5 omhandler repræsentation og beregning af forventede akkumulerede cash-flows i forbindelse med henholdsvis bonus, tilfældig policetageradfærd og dobbeltstokastik. I Kapitel 3, som indeholder manuskriptet *Ahmad, Buchardt og Furrer (2020)*, betragter vi gennemsnitsrentekontrakter og bonusordningen *ydelsesopskrivning*, hvor dividender benyttes til at købe ekstra ydelser. Under forudsætning af at dividendestrategien er affin i antallet af tilkøbte ydelser, udleder vi en procedure til beregning af markedsværdien af bonusbetalinger, der effektivt kombinerer simulation af finansrisiko med klassiske metoder for forsikringsrisiko. Der lægges særlig vægt på tilfældet, hvor antallet af tilkøbte ydelser kun afhænger af finansrisiko, hvorved vi bygger bro mellem kollektive og individuelle synspunkter. Tilfældigt policetageradfærd fører til betalinger skaleret i henhold til tidspunkterne, hvorpå der gøres brug af policetageroptionerne. I Kapitel 4 etablerer vi en forbindelse mellem skaleringsfaktorer og målskift ved at benytte supermartingaler som Radon-Nikodym-afledte. Denne forbindelse kan udnyttes til bekvemt at udlede baglæns- og forlænsmetoder til beregning af prospektive reserver. Kapitel 5, som indeholder artiklen *Buchardt, Furrer og Steffensen (2019)*, studerer forward-overgangsintensiteter i dobbeltstokastiske Markovkædemodeller. Vi etablerer en teoretisk ramme, foreslår et nyt koncept og sammenligner det med tidligere forslag i litteraturen: marginale og tilstandsvise forward-overgangsintensiteter.

Modsat de øvrige undersøgelser fokuserer Kapitel 6, som indeholder artiklen *Furrer (2019)*, på statistiske fremfor sandsynlighedsteoretiske aspekter af flertilstandsmodellering, nemlig erfaringstarifering for flertilstandslivsforsikring. Til dette formål anvender vi empirisk-Bayes metoder på udvidelser af klassiske Markovkædemodeller gennem tilføjelse af flerdimensionel skrøbelighed. Der lægges særlig vægt på tilfældet med gensidigt uafhængige og Gamma-fordelte gruppeeffekter, hvor den klassiske forbindelse til Poisson-regressioner erstattes af en forbindelse til flerdimensionelle negativ-binomial-regressioner. Metoderne og resultaterne illustreres gennem et numerisk eksempel for invalideforsikring med simuleret data.

Skat på investeringsafkast fører til forsikringsbetalinger, som afhænger af investeringsstrategien. Følgelig kan klassiske metoder for markedskonsistent værdiansættelse ikke finde anvendelse. I Kapitel 7, som indeholder artiklen *Buchardt, Furrer og Møller (2020)*, betragter vi kvadratisk afdækning af forsikringsbetalingsprocesser under hensyntagen til skat og omkostninger. Kapitlet adskiller sig fra de øvrige undersøgelser ved ikke at have flertilstandsmodellering som sit centrale tema. Vi foreslår kriteriet skat- og omkostningsmodificeret risikominimering, der tager hensyn til effekten af skatter og omkostninger på tidsværdien af penge. Vi etablerer eksistens og unikhed af en optimal investeringsstrategi relateret til Galtchouk-Kunita-Watanabe-dekompositionen af den intrinsiske værdiproces knyttet til en skat- og omkostningsmodificeret betalingsproces. Undersøgelsen afsluttes med an anvendelse af skat- og omkostningsmodificeret risikominimering rettet mod flertilstandslivsforsikring.

Contents

Chapter 1

Introduction

This thesis contains a series of investigations that are primarily concerned with the conceptual and computational challenges arising from multi-state modeling in actuarial mathematics, specifically within the mathematics of life insurance. They are preluded by this introductory chapter of the following structure. In Sections 1.1–1.2, some aspects of the mathematics of multi-state life insurance, including its interplay with point process theory, are introduced and discussed. The purpose is neither to give a full account of the historical development of the field nor to provide an exhaustive list of recent contributions, but instead to describe in various degrees of detail methods and results of the field that are of particular importance to the investigations of the thesis. Building on this, Section 1.3 concludes with an overview of the main contributions of the thesis and their interrelations. Although the presentation is targeted at actuarial mathematicians with a strong foundation in stochastic process theory, we focus on conceptual contributions by keeping technical aspects to a bare minimum.

1.1 Background

The earliest attempt at a unification of the theory on multi-state modeling for life insurance appears to be Hoem (1969), where the state of the insured is governed by a Markovian jump process. Half a century later, Markov chains remain popular in theory and practice: They are easy to grasp and interpret and computationally viable. An excellent example of the popularity of Markov chains in practice are the Danish risk tables of 1982 (G82), cf. Henriksen et al. (2014) and Gad and Nielsen (2016). Markov chain models are introduced and studied in Subsection 1.1.1 following along the lines of Hoem (1969), Norberg (1991), and Buchardt and Møller (2015).

In recent decades, the inclusion of duration effects – also in relation to policyholder behavior – has received significant interest. Duration effects are relevant since the infinitesimal probabilities of insurance events are duration dependent (think: probability of recovery from disability in regards to time since onset of disability) and since insurance contracts are designed in such a way that payments could depend on e.g. the time of disability or the retirement age of the insured. The inclusion of duration effects and other extensions of Markov chain models are discussed in Subsection 1.1.2.

The interest in models for which the infinitesimal jump probabilities and payments are also allowed to depend on the duration since the last jump actually predates Hoem (1969), see Janssen (1966). Early contributions to the field are methodologically very different from the modern approach; the latter relies on martingale methods for marked point processes and multivariate counting processes. The links between actuarial multi-state modeling and point process theory, while already established and utilized in Hoem and Aalen (1978), were highlighted in a series of papers by Ragnar Norberg in the early 90s, see Norberg (1990, 1991, 1992). These and other aspects of multi-state modeling are discussed in more detail in Subsection 1.2.1.

1.1.1 Markov chain models

In Markov chain models, the state of the insured is governed by a Markov chain $Z = (Z_t)_{t\geq 0}$ on a finite state space $\mathcal{J}$. The elements of $\mathcal{J}$ represent biometric and behavioral states of the insured related to e.g. disability and retirement. The filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ contains exactly the information generated by Z. It constitutes the available information. The chain is assumed to admit suitably regular transition rates μ such that the transition probabilities p satisfy Kolmogorov's classic backward and forward differential equations:

$$
\begin{aligned}
&\frac{\partial}{\partial t} p_{ij}(t,s) = \sum_{\ell\in\mathcal{J}:\ell\neq i} \mu_{i\ell}(t) p_{ij}(t,s) - \sum_{\ell\in\mathcal{J}:\ell\neq i} \mu_{i\ell}(t) p_{\ell j}(t,s), \\
&\frac{\partial}{\partial s} p_{ij}(t,s) = -p_{ij}(t,s) \sum_{\ell\in\mathcal{J}:\ell\neq j} \mu_{j\ell}(s) + \sum_{\ell\in\mathcal{J}:\ell\neq j} p_{i\ell}(t,s)\mu_{\ell j}(s), \\
&p_{ij}(t,t) = \mathbb{1}_{\{i=j\}}.
\end{aligned}
$$

Denote by N the multivariate counting process associated with Z. Its components N_{jk} are given by

$$
N_{jk}(t) = \#\left\{s \in (0,t] : Z_{s-} = j, Z_s = k\right\}.
$$

The insurance contract is modeled by a payment process B describing the accumulated benefits less premiums. It is assumed to take the form

$$
B(\mathrm{d}t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_t=j\}} b_j(t)\,\mathrm{d}t + \sum_{j,k\in\mathcal{J}:k\neq j} b_{jk}(t) N_{jk}(\mathrm{d}t) \tag{1.1.1}
$$

for suitably regular deterministic sojourn payment rates b_j and transition payments b_{jk}. The payments taking this form is a key assumption.

Regarding the inclusion of lump sum payments and the inclusion of point probability mass in the distribution of the jumps of the chain, we refer to the more technical presentation of Milbrodt and Stracke (1997). It also appears very fruitful for these endeavors to take as the starting point the product-integral, see Gill and Johansen (1990, Section 4.4). In practice, lump sum payments and point probability mass might be dealt with on a case to case basis.

The time value of money is described by a savings account S_0, which we suppose admits a suitably regular short rate r such that

$$S_0(\mathrm{d}t) = r(t)\, S_0(t)\, \mathrm{d}t.$$

In this section, financial risk is disregarded by assuming r to be deterministic. The interplay with financial mathematics is discussed in more detail in Subsection 1.2.2.

The present value PV of all future payments is given by

$$PV(t) = \int_t^\infty e^{-\int_t^s r(u)\,\mathrm{d}u}\, B(\mathrm{d}s).$$

By diversifying (averaging out unsystematic insurance risk, think: law of large numbers in connection with a sizeable portfolio), we arrive at the prospective reserve V given by

$$V(t) = \mathbb{E}[\,PV(t)\,|\,\mathcal{F}_t] = \int_t^\infty e^{-\int_t^s r(u)\,\mathrm{d}u}\, A(t, \mathrm{d}s),$$

where A are so-called expected accumulated cash flows given by

$$A(t, s) = \mathbb{E}[\,B(s) - B(t)\,|\,\mathcal{F}_t].$$

These quantities are of interest not only for valuation, but also for risk management in general and asset liability management specifically.

Note that in the setup of this subsection,

$$A(t, \mathrm{d}s) = \sum_{j \in \mathcal{J}} p_{Z_t j}(t, s) \left(b_j(s) + \sum_{k \in \mathcal{J} : k \neq j} \mu_{jk}(s) b_{jk}(s) \right) \mathrm{d}s.$$

This is a consequence of the non-trivial fact that the differences $t \mapsto N_{jk}(t) - \int_0^t \mathbb{1}_{\{Z_{s-} = j\}} \mu_{jk}(s)\, \mathrm{d}s$ define martingales, see also Subsection 1.2.1. Informally, one might also argue along the lines of $\mathbb{E}[N_{jk}(\mathrm{d}s)\,|\,\mathcal{F}_{s-}] = \mathbb{1}_{\{Z_{s-} = j\}} \mu_{jk}(s)\, \mathrm{d}s$; the notion of martingales rigorizes this way of thinking.

This specific representation of the expected accumulated cash flows allows us to define so-called state-wise expected accumulated cash flows $(A_i)_{i \in \mathcal{J}}$ and state-wise prospective reserves $(V_i)_{i\in\mathcal{J}}$ via

$$A_i(t, \mathrm{d}s) = \sum_{j\in\mathcal{J}} p_{ij}(t,s)\left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j} \mu_{jk}(s)b_{jk}(s)\right)\mathrm{d}s, \tag{1.1.2}$$

$$V_i(t) = \int_t^\infty e^{-\int_t^s r(u)\,\mathrm{d}u}\, A_i(t, \mathrm{d}s). \tag{1.1.3}$$

Since $A(t,\cdot) = A_{Z_t}(t,\cdot)$ and $V(t) = V_{Z_t}(t)$, smart computation of state-wise prospective reserves and/or state-wise expected accumulated cash flows is important to practitioners. This aspect is examined more closely below.

Backward and forward methods

We now assume the existence of a maximal contract time $\eta < \infty$ in the sense that $b_j(t) = 0$ and $b_{jk}(t) = 0$ for $t > \eta$.

From (1.1.2)–(1.1.3) we see that we can compute $V(t_0)$ by first calculating the relevant state-wise expected cash flow, using Kolmogorov's classic forward differential equations to compute the transition probabilities $p(t_0,\cdot)$, and then discounting and accumulating it. The concept of forward method refers exactly to this computational scheme.

The following differential equations can be derived by differentiating (1.1.3) and using Kolmogorov's backward differential equation, cf. Hoem (1969). It generalizes Thiele's differential equation for a term life insurance dating back to 1875 to multi-state life insurance payments. The differential equations read

$$\frac{\mathrm{d}}{\mathrm{d}t}V_i(t) = r(t)V_i(t) - b_i(t) - \sum_{j\in\mathcal{J}:j\neq i} \left(b_{ij}(t) + V_j(t) - V_i(t)\right)\mu_{ij}(t) \tag{1.1.4}$$

with boundary conditions $V_i(\eta) = 0$, $i \in \mathcal{J}$. They are known in the literature as *Thiele's differential equations*. An application of these differential equations yields the state-wise prospective reserves not only for a fixed time point t_0 but for all time points. They provide an alternative to the forward method and, essentially, generalize Kolmogorov's classic backward differential equations. The application of Thiele's differential equations to compute the state-wise prospective reserves for all time points between a fixed initial time point t_0 and the maximal contract time η is referred to as the backward method.

If one intends to compute the prospective reserve at a fixed time point for various short rates, then the forward method is attractive. If one intends to compute the prospective reserve for a fixed short rate for all time points, then the backward method is attractive. Which method you should use thus depends on the nature of the questions you are investigating.

1.1.2 Duration dependence and double stochasticity

While prevalent in practice and computationally rather simple, Markov chain models are unable to fully capture the inherent complexity of multi-state life insurance. In this subsection we discuss some extensions intended to rectify the situation. The extensions relate to the inclusion of duration effects in the form of semi-Markovian models (see e.g. Hoem, 1972; Helwich, 2008; Christiansen, 2012; Buchardt, Møller, and Schmidt, 2015) and policyholder behavior (see e.g. Buchardt and Møller, 2015; Buchardt, Møller, and Schmidt, 2015) and the inclusions of systematic insurance risk via doubly stochastic modeling (see e.g. Christiansen, 2006; Buchardt, 2017).

Semi-Markov models

Define the duration process $U = (U_t)_{t\geq 0}$ by

$$U_t = t - \sup\{s \in [0,t] : Z_s \neq Z_t\}.$$

This process measures the time spent by the insured in its current state. In semi-Markov modeling, the setup presented in Subsection 1.1.1 is extended in two directions. The bivariate process (Z, U) instead of Z itself is required to be Markovian; we say that Z is a semi-Markovian process. Furthermore, the payment process takes the form

$$B(\mathrm{d}t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_t=j\}} b_j(t, U_t)\,\mathrm{d}t + \sum_{j,k\in\mathcal{J}:k\neq j} b_{jk}(t, U_{t-}) N_{jk}(\mathrm{d}t)$$

for suitably regular duration-dependent sojourn payment rates b_j and transition payments b_{jk}. If Z is Markovian <u>and</u> the sojourn payment rates and transition payments are not duration dependent, then we recover the setup of Subsection 1.1.1.

We disregard lump sum payments and point probability mass in the distribution of the jumps. The latter simplification is equivalent to assuming that the compensators of the multivariate counting process are absolutely continuous with respect to the Lebesgue measure. This is obtained by requiring the existence of suitably regular duration-dependent transition rates μ such that the differences

$$t \mapsto N_{jk}(t) - \int_0^t \mathbb{1}_{\{Z_{s-}=j\}} \mu_{jk}(s, U_{s-})\,\mathrm{d}s$$

define martingales. This characterizes the distribution of Z, cf. Subsection 1.2.1. Inclusion of lump sum payments and point probability mass in the distribution of the jumps is discussed in Helwich (2008), which builds on the aforementioned presentation of Milbrodt and Stracke (1997).

The expected accumulated cash flows A are given by

$$A(t, \mathrm{d}s) = \sum_{j\in\mathcal{J}} \int_0^{U_t+s-t} p_{Z_t j}(t, s, U_t, \mathrm{d}z) \left(b_j(s,z) + \sum_{k\in\mathcal{J}:k\neq j} \mu_{jk}(s,z) b_{jk}(s,z) \right) \mathrm{d}s,$$

where p are the transition probabilities defined via

$$p_{ij}(t,s,u,z) = \mathbb{P}(Z_s = j, U_s \leq z \mid Z_t = i, U_t = u).$$

We retain the assumption of a maximal contract time $\eta < \infty$ in the sense that $b_j(t,\cdot) = 0$ and $b_{jk}(t,\cdot) = 0$ for $t > \eta$. The backward and forward methods for Markov chain models can be generalized to semi-Markov models. Since the transition probabilities satisfy implementable forward integro-differential equations (see Buchardt, Møller, and Schmidt, 2015, Section 3), the above results guarantee that the formulation of a forward method is quite straightforward. In regards to the backward method, introduce auxiliary functions $W_i(\cdot,v)$ via

$$\begin{aligned} &W_i(t,v) \\ &= \int_t^\eta e^{-\int_t^s r(u)\,\mathrm{d}u} \sum_{j\in\mathcal{J}} \int_0^{s-v} p_{ij}(t,s,t-v,\mathrm{d}z) \left(b_j(s,z) + \sum_{k\in\mathcal{J}:k\neq j} \mu_{jk}(s,z) b_{jk}(s,z) \right) \mathrm{d}s \end{aligned}$$

for $0 \leq v \leq t \leq \eta$, $i \in \mathcal{J}$. Note that $V(t) = W_{Z_t}(t, t-U_t)$. Since the family of functions with elements $W_i(\cdot,s)$ satisfy implementable backward differential equations (see Adékambi and Christiansen, 2017, Corollary 7.8 and Section 8 with $m = 1$), the formulation of a backward method is also quite straightforward.

The computational schemes presented for the aforementioned Markov chain models are to be executed on a suitable grid of $\mathcal{J} \times [0,\eta]$. In comparison, the computational schemes for semi-Markov models are to be executed on a suitable grid of $\mathcal{J} \times \{(t,s) \in [0,\eta]^2 : s \leq t\}$. This constitutes a significant increase in numerical complexity.

Incidental policyholder behavior

The inclusion of incidental policyholder behavior goes beyond Markov chain and semi-Markov modeling since it introduces additional duration effects: The free policy option (see Henriksen et al., 2014; Buchardt, Møller, and Schmidt, 2015; Buchardt and Møller, 2015; Asmussen and Steffensen, 2020) and the option to retire earlier or later (see Gad and Nielsen, 2016) lead to payments that are scaled by a factor depending on the exercise time(s) of the option(s).

In the following, we suppose that Z is a Markov chain which admits transition rates μ. Methods and results in the case of a semi-Markovian Z can be found in Buchardt, Møller, and Schmidt (2015), while further generalizations are discussed in Subsection 1.3.2.

Let $\mathcal{J} = \mathcal{J}_0 \cup \mathcal{J}_1$ and suppose that $Z_t \in \mathcal{J}_1$ implies $Z_s \in \mathcal{J}_1$ for all $s > t$. Denote by τ the first hitting time of $\mathcal{J}_1$. We interpret $\mathcal{J}_0$ as the possible states of the insured before exercise of the option, $\mathcal{J}_1$ as the possible states of the insured after exercise of the option, and τ as the exercise time of the option. Given some suitably regular scaling factor $\rho \in (0,1]$, our interest lies in payments of the form $B^\rho(\mathrm{d}t) = \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau \leq t\}}} B(\mathrm{d}t)$ with B given by (1.1.1) and maximal contract time η. Thus the payments in states $\mathcal{J}_1$ are now scaled by some factor ρ according to the exercise time of the option (and from which state the insured exercised the option).

The prospective reserve V reads

$$V(t) = \int_t^\eta e^{-\int_t^s r(u)\,\mathrm{d}u} \, A^\rho(t, \mathrm{d}s),$$

where the expected accumulated cash flows A^ρ are given by

$$A^\rho(t,s) = \mathbb{E}\left[\int_t^s \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{t<\tau\leq u\}}} B(\mathrm{d}u) \,\middle|\, Z_t\right] \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau\leq t\}}}. \tag{1.1.5}$$

Let Z^ρ be another Markov chain with values in $\nabla \cup \mathcal{J}$ admitting transition rates μ^ρ of the form

$$\begin{aligned}
\mu^\rho_{jk}(t) &= \rho(t,j)\mu_{jk}(t), && j \in \mathcal{J}_0, k \in \mathcal{J}_1,\\
\mu^\rho_{j\nabla}(t) &= (1-\rho(t,j)) \sum_{k\in\mathcal{J}_1} \mu_{jk}(t), && j \in \mathcal{J}_0,\\
\mu^\rho_{j\nabla}(t) &= 0, && j \in \mathcal{J}_1,\\
\mu^\rho_{\nabla k}(t) &= 0, && k \in \mathcal{J},\\
\mu^\rho_{jk}(t) &= \mu_{jk}(t), && \text{otherwise.}
\end{aligned}$$

Denote by p^ρ the transition probabilities of Z^ρ. Then

$$A^\rho(t,s) = \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau\leq t\}}} \sum_{j\in\mathcal{J}} p^\rho_{Z_t j}(t,s) \left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j} \mu^\rho_{jk}(s) b_{jk}(s) \right) \mathrm{d}s. \tag{1.1.6}$$

This follows by the general results developed in Chapter 4, cf. Subsection 1.3.2. The result can also be established by non-trivial yet straightforward calculations following along the lines of Buchardt and Møller (2015, Appendix A). Indeed, the forward differential equations for Z^ρ are directly comparable to the ρ-modified forward differential equations of Buchardt and Møller (2015, Proposition 6).

In combination, (1.1.5) and (1.1.6) lead to a forward method of the same numerical complexity as in Markov chain models without incidental policyholder behavior since computation of p^ρ is not significantly more involved than computation of p. Concerning the backward method, introduce auxiliary functions $(W^\rho_i)_{i\in\mathcal{J}}$ via

$$W^\rho_i(t) = \int_t^n e^{-\int_t^s r(u)\,\mathrm{d}u} \sum_{j\in\mathcal{J}} p^\rho_{ij}(t,s) \left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j} \mu^\rho_{jk}(s) b_{jk}(s) \right) \mathrm{d}s.$$

Note that $V(t) = W^\rho_{Z_t}(t)\rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau\leq t\}}}$. Since the auxiliary functions $(W^\rho_i)_{i\in\mathcal{J}}$ are state-wise prospective reserves of Z^ρ, pointing to (1.1.4) immediately yields a backward method of the same numerical complexity as in Markov chain models without incidental policyholder behavior. Similar considerations are found in Asmussen and Steffensen (2020, Chapter VII.8).

Doubly stochastic models

In Markov chain and semi-Markov models, the transition rates are assumed known. In fact, they have to be estimated and forecasted; this is closely related to the notion of systematic insurance risk. Recently, doubly stochastic Markov chain models have seen a rise in popularity since they allow one to model systematic insurance risk in a multi-state context, see e.g. Steffensen (2000), Christiansen (2006), Norberg (2013), Buchardt (2014), Biagini, Groll, and Widenmann (2016), and Buchardt (2017).

If the transition rates themselves are suitably regular diffusion processes – in particular, (Z, μ) is then Markovian – the backward and forward methods from Markov chain models generalize as follows. In place of the backward differential equations from (1.1.4), one has backward partial differential equations (see Steffensen, 2000), while in place of Kolmogorov's classic forward differential equations, one has forward partial integro-differential equations (see Buchardt, 2017); this is a rather direct consequence of the Markovianity of the multivariate process (μ, Z) and (unstated) regularity conditions pertaining to smoothness. We conclude that the introduction of stochastic transition rates appears to result in a significant increase in numerical complexity, since solving partial differential equations is considerably more demanding than solving ordinary differential equations.

Despite the discouragement expressed in Norberg (2010), some effort has been put into extending the concept of forward mortality (see e.g. Milevsky and Promislow, 2001; Dahl, 2004; Dahl and Møller, 2006; Bauer, Benth, and Kiesel, 2012) to doubly stochastic multi-state models, cf. Christiansen and Niemeyer (2015), Buchardt (2017), and Buchardt, Furrer, and Steffensen (2019). This aspect of doubly stochastic modeling, which also concerns computability, is discussed in more detail in Subsection 1.3.2.

1.2 Foundation and interplay

In the previous section, we focused primarily on Markov chain modeling and the inclusion of duration effects. Point process theory provides a modeling framework that in particular encompasses these aspects of multi-state modeling. Moving to a more general and abstract framework typically simplifies the mathematics – and presently, the point process view is quite popular. In Subsection 1.2.1, we discuss some elements of point process theory in relation to multi-state life insurance. We focus on probabilistic aspects, although a brief introduction to likelihood theory in context of inference and the link to Poisson regressions is also provided.

The presentation of Section 1.1 disregards financial risk, especially since the interest rate is assumed to be deterministic. The interplay between insurance and finance is extensive, and the integration of methods from financial and actuarial mathematics is important from a theoretical as well as practical point of view. In

Subsection 1.2.2, we discuss some aspects of financial risk in relation to multi-state life insurance.

1.2.1 Point process theory

The central mathematical object in Section 1.1 was a non-explosive jump process Z with values in a finite state space $\mathcal{J}$. Suppose for simplicity that $Z_0 \equiv z_0 \in \mathcal{J}$. Denote by $(T_n)_{n\in\mathbb{N}}$ the jump times of Z and by $(Y_n)_{n\in\mathbb{N}}$ the marks given by $Y_n = Z_{T_n}$. The process (T, Y) is then a non-explosive marked point process with mark space $\mathcal{J}$. The associated multivariate counting process N has components given by

$$N_{jk}(t) = \sum_{n\in\mathbb{N}} \mathbb{1}_{\{T_n \leq t\}} \mathbb{1}_{\{Y_{n-1}=j, Y_n=k\}},$$

where $Y_0 := z_0$. To avoid dealing with local martingales, we impose the regularity conditions

$$\mathbb{E}[N_{jk}(t)] < \infty, \qquad t \geq 0.$$

The jump process Z, the marked point process (T, Y), and the multivariate counting process N encode the same information. In the following, the filtration $\mathbb{F}$ represents this information. If no assumptions are made regarding the intertemporal dependence structure of Z, it is often more convenient to focus instead on (properties of) the marked point process and the multivariate counting process.

More or less modern classics on probabilistic and statistical aspects of point process theory include Andersen et al. (1993), Last and Brandt (1995), and Jacobsen (2006).

Compensators and martingales

According to the Ionescu-Tulcea theorem, the probabilistic model is specifiable via the (conditional) distributions of the marked point process. The Markov chains admitting transition rates μ of Subsection 1.1.1 are for example obtained by setting

$$\begin{aligned}
(1 - F^n(t)) &:= \mathbb{P}(T_{n+1} > t \,|\, T_1, Y_1, \ldots, T_n, Y_n) \\
&= e^{-\int_{T_n}^t \sum_{\ell\in\mathcal{J}:\ell\neq Y_n} \mu_{Y_n\ell}(s)\,\mathrm{d}s}, \\
G_k^n(t) &:= \mathbb{P}(Y_{n+1} = k \,|\, T_1, Y_1, \ldots, T_n, Y_n, T_{n+1} = t) \\
&= \frac{\mu_{Y_n k}(t)}{\sum_{\ell\in\mathcal{J}:\ell\neq Y_n} \mu_{Y_n\ell}(t)}.
\end{aligned} \tag{1.2.1}$$

Martingale techniques play a central role in the theory and application of point processes. As a consequence of the Doob-Meyer decomposition theorem, there exist predictable processes Λ_{jk} such that the differences $M_{jk} := N_{jk} - \Lambda_{jk}$ are martingales. Versions of these compensators Λ of the multivariate counting process N are explicitly given by

$$\Lambda_{jk}(t) = \sum_{n\in\mathbb{N}_0} \int_0^t \mathbb{1}_{(T_n,T_{n+1}]}(s) \mathbb{1}_{\{Y_n=j\}} \frac{G_k^n(s) F^n(\mathrm{d}s)}{1 - F^n(s-)}, \tag{1.2.2}$$

where we have employed the convention $T_0 \equiv 0$. This formula is due to Jacod (1975). Under the specification (1.2.1), we have that

$$\Lambda_{jk}(t) = \sum_{n\in\mathbb{N}_0} \int_0^t \mathbb{1}_{(T_n,T_{n+1}]}(s) \mathbb{1}_{\{Y_n=j\}} \mu_{Y_n k}(s)\,\mathrm{d}s = \int_0^t \mathbb{1}_{\{Z_{s-}=j\}} \mu_{jk}(s)\,\mathrm{d}s.$$

If the compensator is absolutely continuous with resepect to the Lebesgue measure, then the Radon-Nikodym derivative is denoted *intensity process*. The existence of intensity processes is by (1.2.2) equivalent to the absence of point probability mass in the (conditional) distributions of the jumps. Actually, the compensators characterize the distribution of the multivariate counting process, equivalently, the marked point process and the jump process.

A key result in point process theory is the martingale representation theorem and the explicit characterization of the resulting integrand found in its proof, see e.g. Jacobsen (2006, Theorem 4.6.1). If $X = (X_t)_{t\geq 0}$ is a suitably regular real-valued stochastic process with $\mathcal{F}_t$-measurable differences $X_t - X_0$, $t > 0$, then the theorem and its proof imply that

$$\begin{aligned}
\mathbb{E}[X_t \,|\, \mathcal{F}_t] &= \mathbb{E}[X_0 \,|\, \mathcal{F}_0] + X_t - X_0 \\
&\quad + \sum_{j,k\in\mathcal{J}:k\neq j} \int_0^t \sum_{n\in\mathbb{N}_0} \mathbb{1}_{(T_n,T_{n+1}]}(s)\, h_k^n(s)\, M_{jk}(\mathrm{d}s), \\
h_k^n(t) &= \mathbb{E}[X_{t-} \,|\, T_1, Y_1, \ldots, T_n, Y_n, T_{n+1} = t, Y_{n+1} = k] \\
&\quad - \mathbb{E}[X_{t-} \,|\, T_1, Y_1, \ldots, T_n, Y_n, T_{n+1} > t]\,.
\end{aligned} \tag{1.2.3}$$

Comparable results can be found in e.g. Christiansen and Djehiche (2020).

We now turn our attention to an application of point process theory for multi-state life insurance: the derivation of dynamics of prospective reserves. Restrictions on the intertemporal dependence structure, say Markovianity or semi-Markovianity, are important to practitioners since they ensure computability of expected cash flows and prospective reserves, cf. Section 1.1. On the other hand, general results may reveal what is actually happening behind the scenes. They should also lead to the establishment of model-independent concepts and could encourage the development of new mathematical methods.

In the following, technical details are intentionally omitted. We consider a suitably regular payment process B and assume the existence of a suitably regular predictable process Λ^B such that the difference $B - \Lambda^B$ is a martingale. Let

$$X_t = \int_t^\infty e^{-\int_0^s r(u)\,\mathrm{d}u}\,\Lambda^B(\mathrm{d}s).$$

We may then cast the prospective reserve V via $V(t) = e^{\int_0^t r(u)\,\mathrm{d}u}\,\mathbb{E}[X_t\,|\mathcal{F}_t]$. Following along the lines of the proof of Proposition 3.2 in Christiansen and Djehiche (2020), the explicit martingale representation of (1.2.3) yields the stochastic differential equation

$$V(\mathrm{d}t) = r(t)\,V(t)\,\mathrm{d}t - \Lambda^B(\mathrm{d}t) + \sum_{j,k\in\mathcal{J}:k\neq j} H_{jk}(t)\,M_{jk}(\mathrm{d}t),$$

$$\mathbb{1}_{\{Z_{t-}=j\}}H_{jk}(t) = \mathbb{1}_{\{Z_{t-}=j\}}e^{\int_0^t r(u)\,\mathrm{d}u}\Big(\mathbb{E}[X_t\,|\,\mathcal{F}_{t-}, Z_t = k] - \mathbb{E}[X_t\,|\,\mathcal{F}_{t-}, Z_t = j]\Big).$$

Point process techniques have been applied in Møller (1993) for semi-Markovian jump processes and in Norberg (1992, 1996) in the presence of intensity processes to derive dynamics of (state-wise) prospective reserves. The potent idea of using an explicit martingale representation is due to Marcus C. Christiansen, see also Christiansen and Djehiche (2020), Christiansen (2020), and Christiansen and Furrer (2020).

Likelihoods and Poisson regressions

We conclude the survey on point process theory in relation to multi-state life insurance by introducing relevant likelihoods and discussing the link to Poisson regressions. Poisson regressions find widespread use in actuarial practice, cf. Gschlössl, Schoenmaekers, and Denuit (2011) and Furrer (2019).

Let $\tilde{\mathbb{P}}$ be another probability measure, and suppose that $\mathbb{P}_t \ll \tilde{\mathbb{P}}_t$ for all $t \geq 0$, where $\mathbb{P}_t$ and $\tilde{\mathbb{P}}_t$ denote the restrictions of $\mathbb{P}$ and $\tilde{\mathbb{P}}$, respectively, to $\mathcal{F}_t$. Denote by $\tilde{\Lambda}$ the compensators of N with respect to $\tilde{\mathbb{P}}$, and suppose for notational convenience that the components of both Λ and $\tilde{\Lambda}$ are absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym derivatives λ and $\tilde{\lambda}$, respectively. The likelihood process $\mathcal{L} = (\mathcal{L}_t)_{t\geq 0}$ then reads (see e.g. Jacod, 1975)

$$\mathcal{L}_t := \frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\tilde{\mathbb{P}}_t} = \prod_{j,k\in\mathcal{J}:k\neq j} \frac{\exp\Big\{-\Lambda_{jk}(t) + \int_0^t \log\big(\lambda_{jk}(s)\big)N_{jk}(\mathrm{d}s)\Big\}}{\exp\Big\{-\tilde{\Lambda}_{jk}(t) + \int_0^t \log\big(\tilde{\lambda}_{jk}(s)\big)N_{jk}(\mathrm{d}s)\Big\}}.$$

In statistical applications, the probability measure $\tilde{\mathbb{P}}$ serves as a fixed reference measure and the likelihoods are only needed up to a proportionality factor:

$$\mathcal{L}_t \propto \prod_{j,k\in\mathcal{J}:k\neq j} \exp\Big\{-\Lambda_{jk}(t) + \int_0^t \log\big(\lambda_{jk}(s)\big)N_{jk}(\mathrm{d}s)\Big\}. \tag{1.2.4}$$

The results we have presented here pertain to the so-called canonical or self-exciting case, where the available information – described by the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ – is generated by the multivariate counting process itself. In statistical applications, one must account for (time dependent) covariates and general censorship, filtering, and truncation. Central concepts and results in survival and event history analysis include partial likelihoods and the fact that – under certain conditions – general censorship, filtering, and truncation preserve the form and martingale properties of the (partial) likelihood (see e.g. Andersen et al., 1988). This entails that the discussion below actually is of practical relevance.

As already mentioned, there exists a link between the likelihood of (1.2.4) and certain Poisson regressions. We now illustrate this fact. Suppose for simplicity that Z is a Markov chain admitting transition rates μ. We assume that the transition rates are piecewise constant and right-continuous on some grid $0 = t_0 < t_1 < \ldots < t_n = \eta$ of $[0, \eta]$; this can be viewed as an approximation. The likelihood then reads

$$\mathcal{L}_\eta \propto \prod_{j,k \in \mathcal{J}: k \neq j} \prod_{i=1}^{n} (\mu_{jk}(t_i))^{O^i_{jk}} \exp\{-E^i_j \cdot \mu_{jk}(t_i)\}, \tag{1.2.5}$$

where E^i_j and O^i_{jk} are so-called exposures and occurrences given by

$$E^i_j = \int_{t_{i-1}}^{t_i} \mathbb{1}_{\{Z_{s-}=j\}} \, \mathrm{d}s \quad \text{and} \quad O^i_{jk} = N_{jk}(t_i) - N_{jk}(t_{i-1}).$$

If one assumes independent observations (O^i_{jk}), $j, k \in \mathcal{J}$, $k \neq j$, $i = 1, \ldots, n$, with distributions

$$O^i_{jk} \sim \text{Poisson}\big(E^i_j \cdot \mu_{jk}(t_i)\big),$$

one would also arrive at the likelihood of (1.2.5). This observation that Poisson regressions can be motivated by the form of likelihoods for multivariate counting processes has long been an established fact in survival and event history analysis, see e.g. Aalen, Borgan, and Gjessing (2008, Section 5.2.1). In the context of multi-state life insurance, it has been utilized in Furrer (2019), cf. Subsection 1.3.3.

1.2.2 Financial risk

In the previous (sub)sections, the interest rate was assumed deterministic, which disregards financial risk. Already in 1989, Hans Bühlmann described the emergence of *Actuaries of the Third Kind*, that combine actuarial and financial mathematics (Bühlmann, 1989). Concurrently, researchers paid increasing amounts of attention to questions at the interface of insurance and finance (see Møller, 2002) – a trend which appears to have continued to this very day.

In the greater part of this thesis, financial risk plays at most a secondary role. But while the leading role typically sets the stage, the contributions of the remaining cast are of equal importance. In this subsection, we therefore provide a very brief introduction to some aspects of the interplay between financial mathematics and multi-state life insurance. An early survey on the interplay between insurance and finance is Møller (2002), while specifically in regards to financial risk in life insurance, relevant chapters from the textbooks Møller and Steffensen (2007) and Asmussen and Steffensen (2020) might serve as good introductory reading.

Financial valuation principles

The field of financial mathematics is concerned with the modeling of financial markets in relation to e.g. derivatives pricing and portfolio management. In the following, we focus on aspects related to valuation (pricing). Classic actuarial valuation principles are based on diversification of risks. In financial mathematics, valuation (fair pricing) instead relates to the notion of no *arbitrage*; an arbitrage possibility is a risk-free gain with no initial investment.

The pricing of a contract (or claim) in early financial mathematics involves the identification of a *self-financing* investment strategy that replicates the payout of the contract. An investment strategy is said to be self-financing if the value of the corresponding investment portfolio is always exactly the initially invested amount added trading gains. The no-arbitrage price of the contract is then given by the initial investment since any other price would lead to an arbitrage possibility. A claim is said to be *attainable* if there exists a self-financing strategy that replicates its payoff, and the financial market is said to *complete* if every claim is attainable.

There is a clear link between martingale theory and absence of arbitrage and completeness: Absence of arbitrage relates to the existence of an equivalent probability measure making the discounted price processes martingales (a so-called equivalent martingale measure), while uniqueness of said measure relates to completeness. If the financial market is free of arbitrage and complete, then the so-called risk-neutral pricing formula applies.

The completeness property may cease to hold due to various reasons, one of them being the inclusion of uncertainty which is not generated by the financial market, say insurance risk. The interplay between life insurance and finance is discussed below. There exists a multitude of both general and domain specific approaches to pricing in incomplete markets, including quadratic approaches: mean-variance hedging and (local and global) risk-minimization. In Chapter 7, see also Subsection 1.3.4, we take into account taxes on investment returns and propose a modified (global) risk-minimization criterion. Originally, risk-minimization was introduced by Föllmer and Sondermann (1986) and extended so as to allow for general payment processes in Møller (2001).

Interplay between life insurance and finance

The interplay between life insurance and finance is considerable. Most life insurance contracts are long-term contracts, and, consequently, they are sensitive to changes in the time-value of money. Policyholder behavior might also depend on financial risk; insured might e.g. choose to retire earlier or later depending on the situation on the financial market. Furthermore, insurance payments in themselves may depend on financial risk: While obviously the case for so-called unit-linked (equity-linked) contracts, so-called bonus payments in with-profit (participating) contracts are also affected by financial risk, since e.g. the emergence of surplus depends on trading gains and losses. Introductions to with-profit life insurance and bonus can be found in Ramlau-Hansen (1991) and Norberg (1999).

We conclude this subsection by establishing a more direct link to the methods and results of Section 1.1. Suppose to this end that the insurance payment process B does not depend on financial risk and that insurance risk and financial risk are independent. Then e.g. risk-minimization and mean-variance hedging confirm the following Brennan/Schwartz-type valuation formula:

$$V(t) = \int_t^\infty e^{-\int_t^s f(t,u)\,\mathrm{d}u}\, A(t, \mathrm{d}s). \tag{1.2.6}$$

We have here denoted by f the forward interest rate curves and by A the expected accumulated insurance cash flows. The nomenclature is inspired by Møller (2002), where Brennan and Schwartz (1979a,b) are surveyed and their methods and results are compared to pricing methods for incomplete markets. Brennan/Schwartz-type valuation is essentially a two-step procedure: Insurance risk is first diversified, and then arbitrage-free pricing is applied; conceptually this appears to be in the spirit of the Solvency II regulatory framework (see Article 77 in EIOPA, 2009).

It is quite straightforward to extend the forward and backward methods described in Section 1.1 to the present situation, although for the backward method one must work with auxiliary quantities W defined by

$$W(t_0, t) = \int_t^\infty e^{-\int_t^s f(t_0,u)\,\mathrm{d}u}\, A(t, \mathrm{d}s)$$

and thus satisfying $W(t_0, t_0) = V(t_0)$. This makes the valuation formula (1.2.6) convenient for practitioners.

If the payment process depends on financial risk, e.g. due to bonus in with-profit life insurance, cf. Subsection 1.3.2, or if there is dependence between insurance risk and financial risk, then classic forward and backward methods are not directly applicable.

1.3 Overview and contributions

The thesis consists of an introduction and six additional chapters, with each chapter constituting a stand-alone scientific contribution. This entails significant discrepancies in especially notation across chapters. Before we in the following subsections describe in more detail the content of each chapter, we first discuss similarities and differences between chapters.

In Chapter 2, we derive dynamics of so-called state-wise prospective reserves in the presence of non-monotone information. As our main contribution we present a generalization of the so-called stochastic Thiele equations of Norberg (1992, 1996). Chapters 3–5 are concerned with the representation and computation of expected accumulated cash flows in the presence of bonus, incidental policyholder behavior, and double stochasticity, respectively. In Chapter 6, we discuss experience rating for Markov chain models. Contrary to the remaining chapters, we focus on statistical rather than probabilistic or financial aspects of multi-state modeling. Chapter 7 contains a study of the problem of determining risk-minimizing investment strategies for insurance payment processes in the presence of taxes and expenses. It especially differs from the remaining chapters by not having aspects of multi-state modeling as its focal point.

In both Chapter 2 and Chapter 4, no assumptions are made concerning the intertemporal dependence structure of the jump process governing the state of the insured. The methods and results we derive here are essentially independent of the statistical/probabilistic model (probability measure), and thus they shed light on the universality of certain concepts and structures arising from multi-state modeling in the mathematics of life insurance.

While different in their initial focus, we should also like to stress a single yet important methodological similarity between Chapter 2 and Chapter 7: Martingale representation theorems are utilized to great effect in both investigations.

In Chapter 3 and Chapter 6, ready-to-implement solutions targeted at actuarial practitioners are provided. In comparison, the focus of especially Chapter 2 and Chapter 4 is on the development of model-independent concepts and abstract methodology. Correspondingly, different chapters not only investigate different topics, but the styles of presentation also reflect the intent to address differing audiences.

1.3.1 Dynamics of state-wise prospective reserves in the presence of non-monotone information

Chapter 2, which contains the manuscript *Christiansen and Furrer (2020)*, investigates the dynamics of state-wise prospective reserves in the presence of non-monotone information. Discarding information leads to non-increasing flows of information for which classic martingale theory does not apply. Via the *infinitesimal approach* proposed and developed in Christiansen (2020), we derive stochastic differential equations generalizing the stochastic Thiele equations of Norberg (1992, 1996). Secondarily, we present a careful study of the concept of state-wise prospective reserves, and we study valuation and computation when information is discarded upon and after stochastic retirement. The latter in particular serves as an application of the general theory.

In the following, we give a slightly more detailed account of the setup and results of Chapter 2. We focus on aspects relating directly to the primary contribution: the generalization of Ragnar Norberg's stochastic Thiele equations to allow for non-monotone information. Discussions pertaining to the secondary contributions and null-set gymnastics are intentionally omitted.

The state of the insured is governed by a non-explosive jump process Z with values in a finite set $\mathcal{J}$ and $Z_0 \equiv z_0 \in \mathcal{J}$. No assumptions regarding the intertemporal dependence structure are made: Z is not assumed to be e.g. (semi-)Markovian. The associated multivariate counting process is denoted N.

The available information is described by the sequence of σ-algebras $\mathbb{G} = (\mathcal{G}_t)_{t\geq 0}$ given by

$$\mathcal{G}_t = \sigma(\{\underline{\tau}_i \leq t < \bar{\tau}_i\} \cap \{\zeta_i \in C\} : i \in \mathbb{N}, C \in \mathcal{E}),$$

where $(\underline{\tau}_i)_{i\in\mathbb{N}}$ and $(\bar{\tau}_i)_{i\in\mathbb{N}}$ are sequences of stopping times with respect to Z with $\bar{\tau}_i \geq \underline{\tau}_i$ and $(\zeta_i)_{i\in\mathbb{N}}$ is a sequence of random variables with values in a suitably regular measurable space $(E, \mathcal{E})$; each ζ_i is assumed measurable with resepect to the information generated by $(Z_t)_{0\leq t\leq \underline{\tau}_i}$. In other words, the available information consists of elements ζ_i recorded at time $\underline{\tau}_i$ and discarded at time $\bar{\tau}_i$. Information discarding might e.g. result from legal constraints induced by privacy law.

The sequence $\mathbb{G}$ is in general non-monotone and thus not a filtration. With $(T_i)_{i\in\mathbb{N}}$ the point process corresponding to the jumps of Z, we may however recover the monotone information which Z generates by taking $\underline{\tau}_i = T_i$, $\bar{\tau}_i = \infty$, and $\zeta_i = (T_i, Z_{T_i})$. The extended marked point process $(\underline{\tau}_i, \bar{\tau}_i, \zeta_i)_{i\in\mathbb{N}}$ corresponds to a family of random counting measures $\nu = (\nu_{xy})_{y\neq x}$ counting replacement of information $(\zeta_i)_{i\in x}$ by information $(\zeta_i)_{i\in y}$, $x, y \subset \mathbb{N}$, $|x| < \infty$, $|y| < \infty$, $y \neq x$.

We consider a payment process B of the form

$$B(\mathrm{d}t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_{t-}=j\}} b_j(t)\, m(\mathrm{d}t) + \sum_{j,k\in\mathcal{J}:k\neq j} b_{jk}(t) N_{jk}(\mathrm{d}t),$$

where b_j and b_{jk} are suitably regular predictable processes (with respect to the information generated by Z) and the measure m is a sum of the Lebesgue measure and a countable number of Dirac measures. The sojourn payment rate b is given by $b(t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_{t-}=j\}} b_j(t)$. The present value PV of all future payments is given by

$$PV(t) = \int_t^\infty \frac{S_0(t)}{S_0(u)}\, B(\mathrm{d}u)$$

for a suitably regular deterministic savings account S_0. The prospective reserve under information $\mathbb{G}$ is the optional projection $PV^{\mathbb{G}}$ of PV with respect to $\mathbb{G}$. It satisfies

$$PV^{\mathbb{G}}(t) = \mathbb{E}[PV(t) \,|\, \mathcal{G}_t].$$

Denote by I_x the process indicating if exactly information $(\zeta_i)_{i\in x}$ is available. We study the dynamics of (non-classic) state-wise prospective reserves $PV_x^{\mathbb{G}}$ given by

$$PV_x^{\mathbb{G}}(t) = \frac{\mathbb{E}[I_x(t)PV(t) \,|\, (\zeta_i)_{i\in x}]}{\mathbb{E}[I_x(t) \,|\, (\zeta_i)_{i\in x}]}$$

satisfying (according to Section 4 in Christiansen, 2020)

$$PV^{\mathbb{G}}(t) = \sum_x I_x(t) PV_x^{\mathbb{G}}(t).$$

By applying the explicit infinitesimal martingale theorem (see Christiansen, 2020, Theorem 7.1), we show that the (non-classic) state-wise prospective reserves $(PV_x^{\mathbb{G}})_x$ satisfy the stochastic differential equations

$$\begin{aligned} 0 = \sum_x I_x(t-)\bigg(& PV_x^{\mathbb{G}}(\mathrm{d}t) - PV_x^{\mathbb{G}}(t-)\frac{S_0(\mathrm{d}t)}{S_0(t-)} + b_x^{\mathbb{G}-}(t)\, m(\mathrm{d}t) \\ & + \sum_{y:y\neq x} \int \mathcal{R}^{\mathbb{G}-}(t,x,y,e)\, g_{xy}^{\mathbb{G}-}(\mathrm{d}t\times\mathrm{d}e) \\ & - \sum_{y:y\neq x} \int \mathcal{R}^{\mathbb{G}}(t,y,x,e)\, g_{yx}^{\mathbb{G}}(\mathrm{d}t\times\mathrm{d}e)\bigg). \end{aligned} \tag{1.3.1}$$

The derivation of the stochastic differential equations (1.3.1) constitutes the main contribution of the chapter. The processes $\mathcal{R}^{\mathbb{G}-}$ and $\mathcal{R}^{\mathbb{G}}$ are so-called sums at risk giving the change in prospective reserve at information arrival and discarding, respectively, while $b_x^{\mathbb{G}-}$ is the $\mathbb{G}$-averaged sojourn payment rate in information state x given by

$$b_x^{\mathbb{G}-}(t) = \frac{\mathbb{E}[I_x(t-)b(t) \,|\, (\zeta_i)_{i\in x}]}{\mathbb{E}[I_x(t-) \,|\, (\zeta_i)_{i\in x}]},$$

and the processes $g_{xy}^{\mathbb{G}-}$ and $g_{yx}^{\mathbb{G}}$ are related to the so-called infinitesimal forward and backward compensators of ν with respect to $\mathbb{G}$ via the following informal identities:

$$\begin{aligned} \mathbb{E}[\nu_{xy}(\mathrm{d}t\times\mathrm{d}e) \,|\, \mathcal{G}_{t-}] &= I_x(t-) g_{xy}^{\mathbb{G}-}(\mathrm{d}t\times\mathrm{d}e), \\ \mathbb{E}[\nu_{yx}(\mathrm{d}t\times\mathrm{d}e) \,|\, \mathcal{G}_t] &= I_x(t) g_{yx}^{\mathbb{G}}(\mathrm{d}t\times\mathrm{d}e). \end{aligned}$$

If $\mathbb{G}$ is actually monotone, then the first term is simply the ordinary compensating measure for the random counting measure ν_{xy}, and the second term is simply the random counting measure ν_{yx}.

The last term of (1.3.1) relates to information discarding: Based on the information currently available, the term adjusts the dynamics by taking into account that information discarding might just have occurred. If say $\mathbb{G}$ is actually equal to the monotone information generated by Z, then the last term vanishes and further calculations then allow one to recover the stochastic Thiele equations of Norberg (1992, 1996).

1.3.2 Representation and computation of expected accumulated cash flows

In Chapters 3–5, we study the representation and computation of a range of non-classic expected accumulated cash flows appearing in the presence of bonus, incidental policyholder behavior, and double stochasticity. Discrepancies in especially notation across chapters exist, and each chapter constitutes a stand-alone scientific contribution and should be read as such. Note that in contrast to Chapter 2, our approach here to a larger degree emphasizes forward rather than backward methodology.

Expected accumulated bonus cash flows

Chapter 3, which contains the manuscript *Ahmad, Buchardt, and Furrer (2020)*, is concerned with bonus payments arising in multi-state with-profit life insurance. Contrary to guaranteed payments, bonus payments may depend on financial risk. We consider the bonus scheme known as *additional benefits*, where dividends are used to buy extra benefits. Requiring the dividend strategy to be affine in the number of additional benefits, we derive a procedure for the computation of the market value of bonus payments. Special attention is given to the case where the number of additional benefits only depends on financial risk.

In the following, we give a more detailed account of the setup and results of Chapter 3. The bonus payments B^b take the form $B^b(\mathrm{d}t) = Q(t)B^\dagger(\mathrm{d}t)$, where Q is the number of additional benefits and $B^\dagger$ is a so-called unit bonus cash flow of the form (1.1.1) with maximal contract time $\eta < \infty$. The process Z governing the state of the insured is assumed to be a Markov chain admitting transition rates μ. The dividend yield δ is used as a premium rate to buy $B^\dagger$ on the so-called technical basis, which entails that

$$Q(\mathrm{d}t) = \frac{\delta(t)}{V^{\star,\dagger}_{Z_t}(t)}\,\mathrm{d}t, \qquad Q(0) = 0.$$

Here $(V^{\star,\dagger}_i)_{i\in\mathcal{J}}$ are so-called state-wise technical unit reserves solving Thiele's differential equations with technical transition rates and interest rate $(\mu^\star, r^\star)$.

The time zero market value of bonus payments $V^b(0)$ can be shown to take the form

$$V^b(0) = \mathbb{E}\bigg[\int_0^\eta e^{-\int_0^t r(u)\,\mathrm{d}u}\, A^b(0,\mathrm{d}t)\bigg], \tag{1.3.2}$$

where the expected accumulated bonus cash flow conditionally on financial risk $A^b(0,\cdot)$ reads

$$A^b(0,\mathrm{d}t) = \sum_{j\in\mathcal{J}} p^Q_{Z_0 j}(0,t)\bigg(b^\dagger_j(t) + \sum_{k\in\mathcal{J}:k\neq j}\mu_{jk}(t)\,b^\dagger_{jk}(t)\bigg)\mathrm{d}t \tag{1.3.3}$$

for so-called Q-modified transition probabilities p^Q given by

$$p^Q_{Z_0 j}(0,t) = \mathbb{E}\big[Q(t)\mathbb{1}_{\{Z_t=j\}}\,\big|\,\mathcal{F}^S(t)\big].$$

The financial market is described by the price processes S, and $\mathbb{F}^S = (\mathcal{F}^S_t)_{t\geq 0}$ is the filtration naturally generated by the price processes S.

We consider dividend strategies δ of the form

$$\begin{aligned}\delta(t) = {}&\delta_0\left(t, S_{\cdot\wedge t}, Z_t, \mathcal{I}(t)\right) + \delta_1\left(t, S_{\cdot\wedge t}, Z_t, \mathcal{I}(t)\right)\rho(\tau)^{\mathbb{1}_{(\tau\leq t)}}\\ &+ \delta_2\left(t, S_{\cdot\wedge t}, Z_t, \mathcal{I}(t)\right)Q(t),\end{aligned}$$

where τ is the time of free policy conversion, ρ is the free policy scaling factor, and $\mathcal{I} = (\mathcal{I}(t))_{t\geq 0}$ is the so-called shape of the insurance business consisting essentially of portfolio-wide means describing the performance of the collective.

Using classic techniques, we establish differential and integral equations for the computation of the Q-modified transition probabilities. In combination with the representations (1.3.2)–(1.3.3), this allows us to formulate a forward method for the computation of the time zero market value of bonus payments which efficiently combines simulation of financial risk with classic methods for the outstanding insurance risk. This constitutes the main contribution of the chapter.

We take particular interest in the case with dividend strategies of the form

$$\delta(t) = \tilde{\delta}(t)V^{\star,\dagger}_{Z_t}(t)$$

for some $\mathbb{F}^S$-adapted process $\tilde{\delta}$. In this case, Q itself is $\mathbb{F}$-adapted, thus the expected accumulated bonus cash flow reads

$$\begin{aligned}A^b(0,\mathrm{d}t) &= Q(t)A^\dagger(0,\mathrm{d}t),\\ A^\dagger(0,\mathrm{d}t) &= \sum_{j\in\mathcal{J}} p_{Z_0 j}(0,t)\bigg(b^\dagger_j(t) + \sum_{k\in\mathcal{J}:k\neq j}\mu_{jk}(t)\,b^\dagger_{jk}(t)\bigg)\mathrm{d}t.\end{aligned}$$

This simplifies the numerical procedure and allows us to bridge the conceptual gap between the individual point of view expressed in Bruhn and Lollike (2020) and Falden and Nyegaard (2020) and the collective point of view expressed in Jensen and Schomacker (2015).

Incidental policyholder behavior and change of measure techniques

In Chapter 4, we study the expected accumulated cash flows that arise when payments are scaled by factors depending on the exercise times of options related to incidental policyholder behavior. We relate to the scaling factors a new probability measure allowing for classic representations of the expected accumulated cash flows. The measure is explicitly characterized in terms of the original measure and the scaling factors. Our methods and results generalize earlier approaches in the literature for (semi-)Markov models.

In the following, we give a more detailed account of the setup, methodology, and results of Chapter 4. For the sake of clarity, we consider only the case of a single option. In Chapter 4, a finite number of options are considered.

The payments of interest B^ρ take the form

$$B^\rho(\mathrm{d}t) = \rho(\tau)^{\mathbb{1}_{\{\tau \leq t\}}}\, B(\mathrm{d}t),$$

where B is a suitably regular payment process, τ is the exercise time of the option, and $\rho \in (0,1]$ is the corresponding suitably regular scaling process. The process B is adapted to and ρ is predictable with respect to the information $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ generated by a general non-explosive jump process Z. The jump process takes values in a countable set $\mathcal{J}$ of the form $\mathcal{J} = \{\nabla\} \cup \mathcal{J}_0 \cup \mathcal{J}_1$ with $Z_0 \equiv z_0 \neq \nabla$. No assumptions regarding the intertemporal dependence structure are made: Z is not assumed to be e.g. (semi-)Markovian. The exercise time of the option τ corresponds to the first hitting time of $\mathcal{J}_1$ by Z, i.e. $\tau := \inf\{s \in [0,\infty) : Z_s \in \mathcal{J}_1\}$. We suppose that $\mathbb{P}(\zeta < \infty) = 0$, where $\eta := \inf\{s \in [0,\infty) : Z_s \in \nabla\}$ is the first hitting time of $\{\nabla\}$ by Z, and without loss of generality that $Z_t \in \mathcal{J}_1$ implies $Z_s \in \mathcal{J}_1$ for all $s > t$. This in particular warrants the interpretation of ∇ as artificial: It plays no role under the original measure $\mathbb{P}$.

We are interested in representation and computation of the expected accumulated cash flows $A^\rho(t,s) = \mathbb{E}[B^\rho(s) - B^\rho(t) \,|\, \mathcal{F}_t]$. To this end, we explicitly construct a probability measure $\mathbb{P}^\rho$ ensuring the desirable identity

$$A^\rho(t,s) = \mathbb{E}^\rho\left[\int_t^s \mathbb{1}_{\{\zeta > u\}} B(\mathrm{d}u) \,\middle|\, \mathcal{F}_t\right] \rho(\tau)^{\mathbb{1}_{\{\tau\leq t\}}} \tag{1.3.4}$$

with $\mathbb{E}^\rho$ denoting $\mathbb{P}^\rho$-integration. Denote by N the multivariate counting process associated with Z and by Λ the compensators of N with respect to $\mathbb{P}$. The probability measure $\mathbb{P}^\rho$ can be characterized via the compensators Λ^ρ of N with resepect to $\mathbb{P}^\rho$. We show that they take the form

$$\begin{aligned}
\Lambda^\rho_{jk}(\mathrm{d}t) &= \mathbb{1}_{\{\zeta \geq t\}}\, \rho(t)\, \Lambda_{jk}(\mathrm{d}t), && j \in \mathcal{J}_0, k \in \mathcal{J}_1, \\
\Lambda^\rho_{jk}(\mathrm{d}t) &= \mathbb{1}_{\{\zeta \geq t\}} (1-\rho(t)) \sum_{\ell \in \mathcal{J}_1} \Lambda_{j\ell}(\mathrm{d}t), && j \in \mathcal{J}_0, k = \nabla, \\
\Lambda^\rho_{jk}(t) &= \mathbb{1}_{\{\zeta \geq t\}}\, \Lambda_{jk}(\mathrm{d}t), && \text{otherwise.}
\end{aligned}$$

The result has quite an intuitive interpretation: The probability of receiving payments rather than the payments themselves are scaled – to identical effect in expectation. The result may find use in actuarial practice to conveniently derive backward and forward methods. We exemplified such an application in Subsection 1.1.2.

Since the process $t \mapsto \rho(\tau)^{\mathbb{1}_{\{0<\tau\leq t\}}}$ is actually a $\mathbb{P}$-supermartingale, by (1.3.4) the probability measure $\mathbb{P}^\rho$ is a so-called Föllmer measure for this supermartingale, cf. Perkowski and Ruf (2015). The idea of describing supermartingales as Radon-Nikodym derivatives appears not to have found application in multi-state life insurance hitherto.

Forward transition rates

The generalization of forward mortalities to multi-state models is non-trivial and various definitions have been proposed. In Chapter 5, which contains the paper *Buchardt, Furrer, and Steffensen (2019)*, we establish a theoretical framework for the discussion of forward transition rates in doubly stochastic Markov chain models. We propose a new concept, *forward equations rates*, and compare it to earlier proposals in the literature: so-called marginal and state-wise forward transition rates.

In the following, we give a more detailed account of the framework of Chapter 5. The payment process of interest B is of the form (1.1.1), while the process Z governing the state of the insured is a doubly stochastic Markov chain with suitably regular transition rates μ. This means that the transition rates are stochastic processes, and the available information is represented by the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ generated by (Z, μ).

We are interested in representation and computation of the resulting expected accumulated cash flows A. They can be shown to be given by

$$A(t, \mathrm{d}s) = \sum_{j\in\mathcal{J}} \mathbb{E}\big[\mathbb{1}_{\{Z_s=j\}} \,\big|\, \mathcal{F}_t\big]\, b_j(s)\,\mathrm{d}s + \sum_{j,k\in\mathcal{J}:k\neq j} \mathbb{E}\big[\mathbb{1}_{\{Z_s=j\}}\mu_{jk}(s) \,\big|\, \mathcal{F}_t\big]\, b_{jk}(s)\,\mathrm{d}s.$$

The idea behind forward transition rates is inspired by replacement results for forward mortalities and forward interest rates. For fixed $t \geq 0$ we look for suitably regular $\mathcal{F}_t$-measurable functions $m_{jk}(t,\cdot)$ and $p^m_{Z_t k}(t,\cdot)$, with the latter solving the forward differential equations

$$\begin{aligned} \frac{\partial}{\partial s} p^m_{Z_t j}(t,s) &= p^m_{Z_t j}(t,s) \sum_{k\in\mathcal{J}:k\neq j} m_{jk}(t,s) + \sum_{k\in\mathcal{J}:k\neq j} p^m_{Z_t k}(t,s) m_{kj}(t,s), \\ \sum_{j\in\mathcal{J}} p^m_{Z_t j}(t,s) &= 1, \\ p^m_{Z_t j}(t,t) &= \mathbb{1}_{\{Z_t=j\}}, \end{aligned} \tag{1.3.5}$$

satisfying the identities

$$p^m_{Z_t j}(t,s) = \mathbb{E}\big[\mathbb{1}_{\{Z_s=j\}} \,\big|\, \mathcal{F}_t\big], \tag{1.3.6}$$

$$p^m_{Z_t j}(t,s) m_{jk}(t,s) = \mathbb{E}\big[\mathbb{1}_{\{Z_s=j\}}\mu_{jk}(s) \,\big|\, \mathcal{F}_t\big]. \tag{1.3.7}$$

If (1.3.6)–(1.3.7) hold, then

$$A(t,\mathrm{d}s) = \sum_{j\in\mathcal{J}} p^m_{Z_t j}(t,s)\left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j} m_{jk}(t,s)b_{jk}(s)\right)\mathrm{d}s.$$

In other words, the expected accumulated cash flow $A(t,\cdot)$ may conveniently be computed via the forward differential equations (1.3.5); this mimics the forward method for classic Markov chain models, cf. Subsection 1.1.1. The result is a two-step algorithm, consisting first of a calibration of suitable m and then the application of a classic procedure.

We propose and study a new definition of forward transition rates, so-called *forward equations rates*, defined uniquely (at least for decrement models) by being $\mathcal{F}^\mu_t$-measurable solutions to (1.3.5)–(1.3.6), where the filtration $\mathbb{F}^\mu = (\mathcal{F}^\mu_t)_{t\geq 0}$ describes the information generated solely by μ. They are compared to the marginal forward transition rates from Christiansen and Niemeyer (2015) and the state-wise forward transition rates introduced in Buchardt (2017). We find that different concepts of forward transition rates reveal different aspects of doubly stochastic modeling.

In Norberg (2010) it is stated that a definition ought to be "fruitful in the sense of shedding some light on objects other than the one defined" (Norberg, 2010, p. 111), and it is argued that papers advocating forward mortalities rates fail to deliver in this respect – accordingly, "forward rates need the applications more than the applications need them" (Norberg, 2010, p. 111). I am not opposed to Ragnar Norberg's remarks: Mathematically, forward transition rates are probably not the way forward. However, forward rate thinking might appeal to practitioners due to the aforementioned convenient two-step procedure it gives the impression of providing, and to my knowledge, forward rate thinking (at least unconsciously) remains prevalent in practice. This exposes the need to investigate even more closely the theoretical as well as practical pros and cons of these concepts, and, in my opinion, it therefore supports the relevance of the investigations presented in e.g. Christiansen and Niemeyer (2015) and Buchardt, Furrer, and Steffensen (2019).

1.3.3 Experience rating using an empirical Bayes and multivariate frailty approach

In Chapter 6, which contains the paper *Furrer (2019)*, we discuss experience rating for multi-state life insurance in terms of shrinkage estimation of group effects. Specifically, we apply empirical Bayes methods to a multivariate frailty extension with latent group effects of classic Markov chain models. Special attention is given to the case where the group effects are mutually independent and Gamma-distributed. The classic link to Poisson regression is replaced by a link to multivariate negative binomial regressions, while under quadratic loss shrinkage estimates are given by well-known credibility formulae.

In the following, we give a slightly more detailed account of the setup and results of Chapter 6. The setup consists of independent groups of insured, where conditionally on a collection of latent group effects, the processes governing the states of the insured are independent Markov chains admitting transition rates. These (conditional) transition rates are assumed to take a very specific form: They consist of common base transition rates μ scaled by the latent group effects. The inclusion of latent group effects introduces dependence within groups and heterogeneity between groups.

In the classic setting without latent effects, the product structure of the likelihood, cf. Subsection 1.2.1, is of great importance to practitioners since it – depending on parametrization – enables splits into simpler terms. We study the impact of latent effects on this facet by characterizing model features (relating to parametrization and prior dependence structures) which retain the product structure of relevant likelihood.

Particularly simple shrinkage estimation is obtained by requiring the latent groups effects to be mutually independent with marginal $\Gamma(\psi_i^{-1}, \psi_i^{-1})$-distributions and by assuming that the transition rates are suitably distinctly parameterized and piecewise constant. Utilizing the link to Poisson regressions, cf. Subsection 1.2.1, estimation of the base transition rates μ and the prior variances ψ is then possible via multivariate negative binomial regressions, while empirical Bayes methodology suggests estimating the group effects by the Bayes estimator under e.g. quadratic loss. The latter estimate is shown to satisfy a well-known credibility formula. The investigation concludes with a numerical example for disability insurance using simulated data.

1.3.4 Tax- and expense-modified risk-minimization

Chapter 7, which contains the paper *Buchardt, Furrer, and Møller (2020)*, examines quadratic hedging of insurance payment processes in the presence of taxes and expenses. We propose the criterion of tax- and expense-modified risk-minimization, which takes into account the effect of taxes and expenses on the time value of money. As our main result, we establish existence and uniqueness of an optimal investment strategy related to the Galtchouck-Kunita-Watanabe decomposition of the intrinsic value process associated with a tax- and expense-modified payment process.

In the following, we give a more detailed account of the setup, methods, and results of Chapter 7. The setup consists of an arbitrage-free market with savings account S_0 and additional suitably regular price processes $(S_1, \ldots, S_d)$ with maximal contract time $\eta > 0$. The price processes are modeled under some equivalent martingale measure $\mathbb{Q}$. We assume the savings account admits a suitably regular but possibly stochastic short rate r. The total payment process B^{total} consists of suitably regular insurance payments B as well as tax- and expense payments B^{tax} and B^{e}. As a key modeling assumption the latter take the form

$$B^{\text{tax}}(h, \mathrm{d}t) = \gamma(t-) \sum_{j=0}^{d} h_j(t)\, S_j(\mathrm{d}t) \;\text{ and }\; B^{\text{e}}(h, \mathrm{d}t) = \delta(t)\, V(h,t)\, \mathrm{d}t,$$

where h is an investment strategy, V is the undiscounted value process, $\gamma \in [0,1)$ is a suitably regular tax rate, and δ is a suitably regular expense rate. In other words, taxes are paid continuously at rate γ as a fraction of all returns from the investment strategy, while expenses are paid continuously at rate δ as a fraction of the value of the investment strategy.

Since the total payment process B^{total} depends on the investment strategy, classic (global) risk-minimization is not applicable. We propose a new criterion, namely tax- and expense-modified risk minimization, which differs from classic risk-minimization since a tax- and expense-modified savings account is used as numeraire. An investment strategy $\tilde{h}$ is said to be risk-minimizing in the presence of taxes and expenses if it is 0-admissible, i.e. $V(\tilde{h}, \eta) = 0$, and minimizes for all $t \in [0, \eta]$ the tax- and expense-modified risk process $\tilde{R}$ defined by

$$\tilde{R}(h,t) = \mathbb{E}^{\mathbb{Q}}\left[\left(\tilde{C}(h,\eta) - \tilde{C}(h,t) \right)^2 \middle| \mathcal{F}_t \right],$$

where $\tilde{C}$ is the tax- and expense-modified cost process defined by $\tilde{C}(h, \mathrm{d}t) = \tilde{S}_0^{-1}(t)\, C(h, \mathrm{d}t)$. Here $\tilde{S}_0$ is the modified savings account given by

$$\tilde{S}_0(\mathrm{d}t) = ((1-\gamma(t))r(t) - \delta(t))\, \tilde{S}_0(t)\, \mathrm{d}t,$$

while C is the undiscounted cost process given by

$$C(h, \mathrm{d}t) = V(h, \mathrm{d}t) - \sum_{j=0}^{d} h_j(t)\, S_j(\mathrm{d}t) + B^{\text{total}}(h, \mathrm{d}t).$$

Denote by $\tilde{\mathcal{V}}$ the so-called intrinsic value process associated with the tax- and expense-modified insurance payments. It is the $\mathbb{Q}$-martingale given by

$$\tilde{\mathcal{V}}(t) = \mathbb{E}^{\mathbb{Q}}\left[B(0) + \int_0^{\eta} \tilde{S}_0^{-1}(t)\, B(\mathrm{d}t) \middle| \mathcal{F}_t \right].$$

Using classic techniques we prove the existence and uniqueness of a risk-minimizing investment strategy $\tilde{h}$ in the presence of taxes and expenses. It is given by

$$\tilde{h}_j(t) = \frac{1}{1-\gamma(t-)} e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\, \mathrm{d}u} h_j^{\tilde{\mathcal{V}}}(t),$$

for $j = 1, \dots, d$, where $h^{\tilde{\mathcal{V}}}$ is the integrand appearing in the Galtchouk-Kunita-Watanabe decomposition of $\tilde{\mathcal{V}}$ with rsepect to the discounted price process S_j/S_0, while the amount invested in the savings account $\tilde{h}_0$ is determined such that

$$V(\tilde{h}, t) = \mathbb{E}^{\mathbb{Q}}\left[\int_t^{\eta} e^{-\int_t^s ((1-\gamma(u))r(u)-\delta(u))\,\mathrm{d}u}\, B(\mathrm{d}t) \,\middle|\, \mathcal{F}_t\right].$$

We also argue that tax- and expense-modified risk-minimization is equivalent to an alternative approach of classic risk-minimization for an artificial market consisting of after-tax and after-expense assets, and we show that tax- and expense-modified risk-minimization is consistent with classic risk-minimization in the sense that a subsequent application of risk-minimization confirms the optimal investment strategy. The investigation is concluded with an application of tax- and expense-modified risk minimization to multi-state life insurance.

Chapter 2

Dynamics of state-wise prospective reserves in the presence of non-monotone information

This chapter contains the manuscript *Christiansen and Furrer (2020).*

Abstract

In the presence of monotone information, stochastic Thiele equations describing the dynamics of state-wise prospective reserves are closely related to the classic martingale representation theorem. When the information utilized by the insurer is non-monotone, classic martingale theory does not apply. By taking an infinitesimal approach, we derive generalized stochastic Thiele equations that allow for information discarding. The results and their implication in practice are illustrated via examples where information is discarded upon and after stochastic retirement.

Keywords: Life insurance; Stochastic Thiele equations; Infinitesimal martingales; Marked point processes; Stochastic retirement

2.1 Introduction

Life insurers frequently employ reduced information in the valuation of liabilities due to e.g. legal constraints and data privacy considerations or to achieve model simplifications. The possibility of information discarding leads to potentially decreasing flows of information for which classic martingale theory does not apply. Based on the novel *infinitesimal approach* proposed and developed in Christiansen (2020), we study the dynamics of so-called state-wise prospective reserves in the presence of non-monotone information. Our main contribution is a generalization of the stochastic Thiele equations of Norberg (1992, 1996) to allow for non-monotone information. Secondary contributions include a careful study of the concept of state-wise prospective reserves and a discussion of current actuarial practices regarding valuation in relation to information discarding upon and after stochastic retirement.

In this paper, the only source of randomness consists of the state of the insured, which is modeled as a non-explosive pure jump process on a finite state space. This places our work within the field of multi-state life insurance mathematics. The definitions of retrospective and prospective reserves in Norberg (1991) encompass non-monotone information, and under (semi-)Markovian assumptions specific instances of non-monotone information appear in the study of retrospective reserves and bonus prognosis, see Norberg (1991, 1999, 2001) and Helwich (2008). But to our knowledge, the literature contains no attempts at the development of a unifying theory for non-Markovian models under non-monotone information. Our contribution constitutes the first step towards this goal, since we impose no restrictions on the intertemporal dependency structure and allow for general information discarding occurring at stopping times w.r.t. the state of the insured.

The multi-state approach to life insurance dates back at least to Hoem (1969), where Thiele equations describing the dynamics of the state-wise prospective reserves are derived under the assumption that the process governing the state of the insured is Markovian. These differential/integral equations were revisited by Norberg in his seminal paper Norberg (1991) and have since been generalized in various directions. This includes relaxing the assumption of Markovianity to allow for duration dependency (semi-Markovianity), taking market risks into account, and the study of higher order moments of prospective reserves, see e.g. Møller (1993), Steffensen (2000), Helwich (2008), Adékambi and Christiansen (2017), and Bladt, Asmussen, and Steffensen (2020). We should mention that while the approach of Steffensen (2000) is very general, the results are only established under strict smoothness conditions that might not be satisfied in practice.

The ordinary Thiele equations are essentially Feynman-Kac type results. In contrast, the stochastic Thiele equations of Norberg (1992, 1996) are stochastic differential equations that apply irregardless of the intertemporal dependency structure and reveal the universality of Thiele's original equation. Furthermore, under Markovian assumptions, stochastic Thiele equations can be used to elegantly derive Feynman-Kac formulas for the prospective reserve.

In the presence of monotone information, the dynamics of prospective reserves are characterized by identifying integrands in the classic martingale representation theorem for random counting measures (Norberg, 1992, 1996; Christiansen and Djehiche, 2020). In similar fashion, our approach relies on the infinitesimal martingale representation theorem of Christiansen (2020), which extends the classic martingale representation theorem for random counting measures to allow for non-monotone information. Essentially, our methodology and results accompany Christiansen (2020); while Christiansen (2020) contains the general theory for so-called infinitesimal compensators and infinitesimal martingales, this theory is here applied to multi-state life insurance.

Although we focus on state-wise prospective reserves and their dynamics, we expect the setting and mathematical techniques presented here to be applicable beyond this specific application, e.g. in relation to estimation and efficient computation of expected cash flows and reserves in the presence of non-monotone information. Broadly speaking, with this paper we initialize a program that aims at the development of general mathematical methodology for multi-state life insurance in the presence of non-monotone information.

The paper is structured as follows. In Section 2.2, we present the probabilistic setup and the main examples concerning information discarding upon and after retirement. In Section 2.3, we develop a mathematically sound concept of state-wise prospective reserves in the presence of potentially non-monotone information. Section 2.4 contains our main result, namely a generalization of the stochastic Thiele equations to allow for non-monotone information, and its application to information discarding upon and after retirement. In particular, we illustrate the pertinence and usefulness of the generalized stochastic Thiele equations by deriving Feynman-Kac formulas beyond the (semi-)Markovian case.

2.2 Monotone and non-monotone information structures

In this section, we introduce a general modeling framework for the random pattern of states of the insured in the presence of non-monotone information. The framework is strongly related to the general theory of non-monotone information for jump processes introduced by Christiansen (2020). To clarify the theoretical as well as practical relevance of an approach allowing for non-monotone information and general intertemporal dependency structures, we further discuss a motivating example concerning stochastic retirement. This leads to the specification of some explicit cases of non-monotone information that serve as the main examples in the ensuing investigation.

2.2.1 General setting

Let $(\Omega, \mathcal{A}, P)$ be a complete probability space with null sets $\mathcal{N}$, and let $Z = (Z_t)_{t \geq 0}$ be a random pattern of states (pure jump process) on the finite state space $S = \{1, \dots, J+1, J+2\}$, $J \in \mathbb{N}_0$, with initial state $Z_0 \equiv z_0 \in S$, giving at each time t the state of the insured in S.

The total information available is denoted $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$; it is the right-continuous and complete filtration given by

$$\mathcal{F}_t = \sigma(Z_s : s \leq t) \vee \mathcal{N}.$$

Since $\mathcal{F}$ is a filtration, it represents monotone (increasing) information.

We relate to the random pattern of states Z a multivariate counting process $N = (N(t))_{t\geq 0}$ with components $N_{jk} = (N_{jk}(t))_{t\geq 0}$, $j, k \in S$, $j \neq k$, giving the number of jumps of Z from state j to state k:

$$N_{jk}(t) = \#\{s \in (0, t] : Z_{s-} = j, Z_s = k\}, \qquad t \geq 0.$$

We impose the following technical condition. It ensures that Z is non-explosive and that compensated counting processes are true martingales.

Assumption 2.2.1 (No explosions and true martingales)**.** *We assume that*

$$\mathrm{E}\Big[\sum_{\substack{j,k\in S\\ j\neq k}} N_{jk}(t)\Big] < \infty$$

for all $t \geq 0$.

If we denote by $\mathcal{T}(t)$ the next jump after time t,

$$\begin{aligned} \mathcal{T}(t) &= \inf\{s \in (t, \infty) : Z_s \neq Z_t\}, \\ \mathcal{T}(\infty) &= \infty, \end{aligned}$$

and employ the convention $\inf \emptyset = \infty$, we can also define a marked point process $(\tau_i, Z_{\tau_i})_{i\in\mathbb{N}_0}$ by

$$\tau_0 = 0, \qquad \tau_i = \mathcal{T}(\tau_{i-1}), \quad i \in \mathbb{N},$$

with $Z_\infty = \nabla$ for some arbitrary cemetery state ∇. The marked point process, multivariate counting process, and random pattern of states formulations of the setup are equivalent in the sense that the information generated by these processes agree.

A life insurance contract between the insured and the insurer is stipulated by the specification of a payment process $B = (B(t))_{t\geq 0}$ representing the accumulated benefits minus premiums. In general, we suppose that B is an $\mathcal{F}$-adapted process that has càdlàg sample paths, finite expected variation on compacts (in particular, it has sample paths of finite variation on compacts), and a deterministic initial value $B(0) \in \mathbb{R}$.

2.2.2 Non-monotone information

Due to e.g. legal constraints, privacy considerations, or to achieve model and/or computational simplifications, the insurer might not have access to or desire to utilize all information available to it. Examples include the newly introduced General Data Protection Regulation 2016/679 of the European Union, where Article 17 describes a so-called 'right to erasure', and the restriction to a Markovian type of information even when the Markov property is not satisfied. Representing the resulting utilized information as a subsequence of σ-algebras, one typically finds that the sequence is non-monotone because certain pieces of information are discarded en route.

To describe the information reductions, we introduce a subsequence of σ-algebras as follows. Let $(T_i)_{i\in\mathbb{N}}$ and $(S_i)_{i\in\mathbb{N}}$ be sequences of $\mathcal{F}$-stopping times with $S_i \geq T_i$, $i \in \mathbb{N}$. Further, let $(\zeta_i)_{i\in\mathbb{N}}$ be a sequence of random variables with values in a separable complete metric space E and corresponding Borel σ-algebra $\mathcal{E} := \mathfrak{B}(E)$, and suppose that each ζ_i is $\mathcal{F}_{T_i}$-measurable. For the sake of a convenient notation, without loss of generality we assume that $0 \notin E$. The information ζ_i is recorded at time T_i and then discarded at a later time S_i; here $S_i = \infty$ signifies no discarding. Thus the admissible information at time $t \geq 0$ is given by the σ-algebra $\mathcal{G}_t \subseteq \mathcal{F}_t$ defined by

$$\mathcal{G}_t = \sigma(\{T_i \leq t < S_i\} \cap \{\zeta_i \in A\} : i \in \mathbb{N}, A \in \mathcal{E}) \vee \mathcal{N}, \tag{2.2.1}$$

while the information available immediately before time $t > 0$ is given by the σ-algebra $\mathcal{G}_{t-} \subseteq \mathcal{F}_{t-}$ defined by

$$\mathcal{G}_{t-} = \sigma(\{T_i < t \leq S_i\} \cap \{\zeta_i \in A\} : i \in \mathbb{N}, A \in \mathcal{E}) \vee \mathcal{N}. \tag{2.2.2}$$

We introduce the notation $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ and $\mathcal{G}_- = (\mathcal{G}_{t-})_{t>0}$.

The subsequence of σ-algebras $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ is in general non-monotone and the random times T_i and S_i are not necessarily stopping times w.r.t. $\mathcal{G}$. We do not assume the random times $(T_i)_{i\in\mathbb{N}}$ and $(S_i)_{i\in\mathbb{N}}$ to take a specific order in time other than $T_i \leq S_i$, and we even allow for simultaneous events. We can recover $\mathcal{F}$ by taking $S_i = \infty$, $T_i = \tau_i$, and $\zeta_i = (\tau_i, Z_{T_i})$ for all $i \in \mathbb{N}$, and from this point and onward, that representation is always assumed whenever $\mathcal{G} = \mathcal{F}$.

Let $\mathcal{S} := \{x \subset \mathbb{N} : |x| < \infty\}$ be the finite subsets of the natural numbers. Note that $\mathcal{S}$ is countable. For each $x \in \mathcal{S}$ we define the indicator processes

$$I_x(t) := \begin{cases} 1 & : \bigcap_{i\in x}\{T_i \leq t < S_i\} \cap \bigcap_{i\notin x}(\Omega \setminus \{T_i \leq t < S_i\}), \\ 0 & : \text{ else}, \end{cases}$$

so that $I_x(t)$ is $\mathcal{G}_t$-measurable for each $t \geq 0$ and $x \in \mathcal{S}$. We assume in continuation of Assumption 2.2.1 that

$$\mathrm{E}\left[\sum_{i=1}^{\infty} \mathbf{1}_{\{T_i \leq t\}}\right] < \infty, \quad t \geq 0, \tag{2.2.3}$$

which implies that on each compact interval we can almost surely find at most finitely many random times T_i, S_i, $i \in \mathbb{N}$. As a result, the indicator processes I_x

have càdlàg paths of finite variation on compacts. The family of indicator processes $I := (I_x)_{x\in\mathcal{S}}$ corresponds to the $\mathcal{G}$-adapted non-explosive random pattern of states

$$\mathcal{Z}_t = \sum_{x\in\mathcal{S}} x\, I_x(t), \quad t \geq 0.$$

This random pattern of states describes the state of information: $\mathcal{Z}_t = x$ if and only if exactly the information $(\zeta_i)_{i\in x}$ is available at time t; in particular, the information $(\zeta_i)_{i\notin x}$ has either been recorded and already discarded or is yet to be recorded.

We generally suppose that

$$\sigma(Z_t) \subseteq \mathcal{G}_t, \quad t \geq 0. \tag{2.2.4}$$

Since we assumed that $0 \notin E$, the information at time t and at time $t-$ can be alternatively represented as

$$\begin{aligned} \mathcal{G}_t &= \sigma(\boldsymbol{\zeta}_x I_x(t) : x \in \mathcal{S}) \vee \mathcal{N}, \quad t \geq 0, \\ \mathcal{G}_{t-} &= \sigma(\boldsymbol{\zeta}_x I_x(t-) : x \in \mathcal{S}) \vee \mathcal{N}, \quad t > 0, \end{aligned} \tag{2.2.5}$$

where $\boldsymbol{\zeta}_x := (\zeta_i)_{i\in x}$, $x \in \mathcal{S}$. Let

$$T_{xy} := \inf\{t \geq 0 : I_x(t-)I_y(t) = 1\},$$

using the convention $\inf \emptyset := \infty$. We see that T_{xy} is the exact point in time where the state of information changes from state x to state y by discarding information $\boldsymbol{\zeta}_{x\setminus y}$ and recording information $\boldsymbol{\zeta}_{y\setminus x}$; here we ignore information that is recorded and immediately discarded. The total information either discarded or recorded at time T_{xy} is thus $\boldsymbol{\zeta}_{xy} := (\zeta_i)_{i\in x\Delta y}$, where $x\Delta y = (x \setminus y) \cup (y \setminus x)$.

The extended marked point process $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}}$ corresponds to the random counting measures ν_{xy}, $x, y \in \mathcal{S}$, $y \neq x$, defined as the unique completions of

$$\nu_{xy}([0,t] \times A) := \mathbf{1}_{\{T_{xy}\leq t\}}\mathbf{1}_{\{\boldsymbol{\zeta}_{xy}\in A\}}, \quad t \geq 0, A \in \mathfrak{B}(E_{xy}),$$

where $E_{xy} := E^{|x\Delta y|}$.

If $T_i = \tau_i$ for all $i \in \mathbb{N}$, the σ-algebra $\mathcal{G}_t$ reveals in particular the indices i of the admissible observations and thus gives a lower bound on the number of past discards, cf. Remark 3.1 in Christiansen (2020), which might be an unwanted feature. As further discussed in Christiansen (2020), by considering suitable permutations it is often possible to obtain non-informative indices; in that case, the number of past discards becomes non-admissible information. In the next subsection, we introduce some specific instances of non-monotone information concerning stochastic retirement and embed them into the framework above. In particular, we exemplify how to obtain non-informative indices using suitable permutations.

2.2.3 Stochastic retirement

Suppose $J \geq 1$ and $z_0 \notin \{J+1, J+2\}$, and let δ and η be the first hitting times of $\{J+2\}$ and $\{J+1\}$, respectively, by Z:

$$\begin{aligned}\delta &= \inf\{t \geq 0 : Z_t = J+2\},\\ \eta &= \inf\{t \geq 0 : Z_t = J+1\}.\end{aligned}$$

We think of δ as the time of death and η as the time of retirement. Accordingly, the states $\{1, \ldots, J\}$ describe the health state of the insured up until retirement or death. In this subsection, we assume a decrement structure such that retirement occurs at most once and death is a terminal event:

Assumption 2.2.2 (Decrement structure concerning retirement and death)**.** *We assume that*

$$\begin{aligned}[0,\infty) \ni t &\mapsto \sum_{\substack{j\in S\\ j>J}} \sum_{\substack{k\in S\\ k\leq J}} N_{jk}(t) = 0,\\ [0,\infty) \ni t &\mapsto N_{(J+2)(J+1)}(t) = 0,\end{aligned}$$

almost surely.

Note that the structure of the state space entails that the insurer is not updating its information concerning the health state of the insured upon or after retirement. In Figure 2.1 we have exemplified this setup for the case $J = 2$ corresponding to a disability model allowing for recovery before retirement.

In actuarial practice, it is common to impose some Markovian structure by assuming the random pattern of states Z to be e.g. Markovian or semi-Markovian. In the following, we illustrate why such assumptions might be insufficient and, as

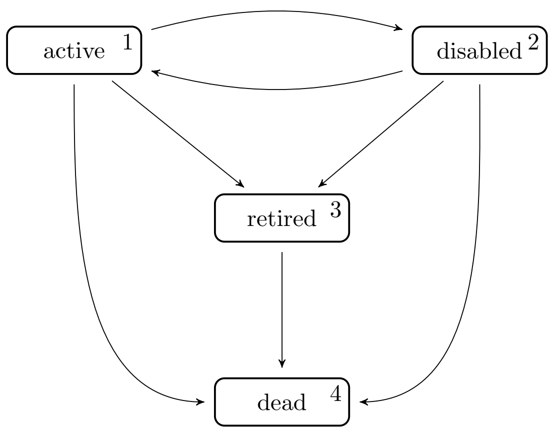

Figure 2.1: *Extension of classic disability model with recovery to allow for stochastic retirement.*

an alternative, how to represent similar assumptions as non-monotone information substructures. This motivates the general non-Markovian framework with non-monotone information introduced in Subsections 2.2.1–2.2.2.

It is natural to imagine the random pattern of states Z as embedded into a larger framework. Let $\tilde{Z}$ be a random pattern of states on an extended state space $\tilde{S} = \{1, \dots, J+1, J+2, \dots, 2J+1\}$ with initial state $\tilde{Z}_0 = \tilde{z}_0 \in \{1, \dots, J\}$. Denote the corresponding multivariate counting process by $\tilde{N}$. Suppose that

$$\mathrm{E}\Big[\sum_{\substack{j,k \in \tilde{S} \\ y \neq x}} \tilde{N}_{jk}(t) \Big] < \infty \tag{2.2.6}$$

for all $t \geq 0$, and that

$$\begin{aligned} [0, \infty) \ni t &\mapsto \sum_{\substack{j \in \tilde{S} \\ j > J}} \sum_{\substack{k \in \tilde{S} \\ k \leq J}} \tilde{N}_{jk}(t) = 0, \\ [0, \infty) \ni t &\mapsto \sum_{\substack{k \in \tilde{S} \\ J < k \leq 2J}} \tilde{N}_{(2J+1)k}(t) = 0, \end{aligned}$$

almost surely. We think of the states $\{J+1, \dots, 2J\}$ as providing information concerning the health state of the insured upon or after retirement. In Figure 2.2 we have exemplified this setup for the case $J = 2$ corresponding to a disability model allowing for recovery and stochastic retirement. In general, we can now redefine Z by

$$Z_t = \begin{cases} \tilde{Z}_t & \text{if } \tilde{Z}_t \in \{1, \dots, J\}, \\ J+1 & \text{if } \tilde{Z}_t \in \{J+1, \dots, 2J\}, \\ J+2 & \text{if } \tilde{Z}_t = 2J+1 \end{cases}$$

for all $t \geq 0$, when we find that $z_0 \notin \{J+1, J+2\}$ and that Assumptions 2.2.1–2.2.2 remain satisfied.

The information available to the insured is represented by the filtration $\tilde{\mathcal{F}} = (\tilde{\mathcal{F}}_t)_{t \geq 0}$ given by

$$\tilde{\mathcal{F}}_t = \sigma(\tilde{Z}_s : s \leq t) \vee \mathcal{N}.$$

In many cases, the information $\tilde{\mathcal{F}}$ is not available to the insurer, and then the insurer must resort to the information given by $\mathcal{F}$; this can e.g. be the case if upon retirement, disability coverage ceases.

It appears consistent with actuarial practice to propose that the underlying random pattern of states $\tilde{Z}$ is Markovian or semi-Markovian. We now study the resulting implications on Z, which is the natural modeling object given information

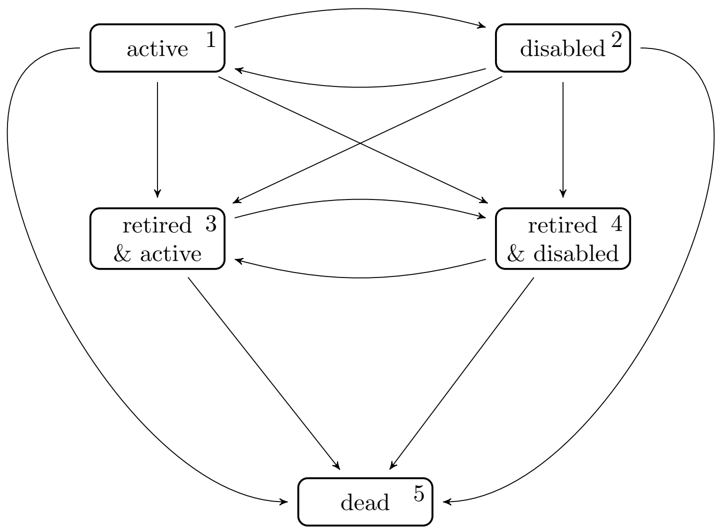

Figure 2.2: *Extension of the disability model with retirement of Figure 2.1 where the health status of the insured remains observed upon and after retirement.*

$\mathcal{F}$. Let $U = (U_t)_{t\geq 0}$ and $\tilde{U} = (\tilde{U}_t)_{t\geq 0}$ be the duration processes associated with Z and $\tilde{Z}$, respectively, given by

$$\begin{aligned} U_t &= t - \sup\{s \in [0,t] : Z_s \neq Z_t\}, \\ \tilde{U}_t &= t - \sup\{s \in [0,t] : \tilde{Z}_s \neq \tilde{Z}_t\}. \end{aligned}$$

Note that $\mathbf{1}_{\{t\leq\eta\}} U_t = \mathbf{1}_{\{t\leq\eta\}} \tilde{U}_t$. Let $U^r = (U^r_t)_{t\geq 0}$ be the time since retirement given by

$$U^r_t = \begin{cases} 0 & \text{if } t < \eta, \\ t - \eta & \text{if } t \geq \eta, \end{cases}$$

let $H = (H_t)_{t\geq 0}$ be the state of the insured just before retirement given by

$$H_t = \begin{cases} Z_t & \text{if } t < \eta, \\ Z_{\eta-} & \text{if } t \geq \eta, \end{cases}$$

and let $U^h = (U^h_t)_{t\geq 0}$ be the duration of the latest sojourn before retirement given by

$$U^h_t = \begin{cases} U_t & \text{if } t < \eta, \\ U_{\eta-} & \text{if } t \geq \eta. \end{cases}$$

Proposition 2.2.3. *Suppose $(\tilde{Z}, \tilde{U})$ is Markovian. Then (Z, U^h, U^r, H) is Markovian. Suppose further that $\tilde{Z}$ is Markovian. Then (Z, U^r, H) is Markovian.*

Proof. See Appendix 2.A. □

It is possible to derive necessary and sufficient conditions for which (semi-)Markovianity of $\tilde{Z}$ implies (semi-)Markovianity of Z, see e.g. Serfozo (1971). In general, these conditions are very restrictive and do not apply to models of actuarial relevance: in this sense, the complex intertemporal dependency structure implied by Proposition 2.2.3 must be taken into account. This serves as a motivation for the general non-Markovian framework presented in Subsection 2.2.1.

Although Proposition 2.2.3 indicates that the mortality as retiree might depend on the past through e.g. the time since retirement and the last health state before retirement, it is common in actuarial practice to rely on a standard mortality table – an example is the longevity benchmark of the Danish financial supervisory authority, cf. Jarner and Møller (2015). This in a sense corresponds to imposing an 'as if' Markovian assumption or, alternatively, to only utilize information corresponding to a specific subsequence of σ-algebras rather than $\mathcal{F}$ itself. Therefore, we introduce two subsequences of σ-algebras $\mathcal{G}^1 = (\mathcal{G}^1_t)_{t\geq 0}$ and $\mathcal{G}^2 = (\mathcal{G}^2_t)_{t\geq 0}$ given by

$$\begin{aligned}
\mathcal{G}^1_t &= \sigma(Z_s \mathbf{1}_{\{Z_t \in \{1,\ldots,J\}\}}, \mathbf{1}_{\{\eta \leq s\}}, \mathbf{1}_{\{\delta \leq s\}} : s \leq t) \vee \mathcal{N}, \\
\mathcal{G}^2_t &= \sigma(Z_s \mathbf{1}_{\{Z_t \in \{1,\ldots,J\}\}}, \mathbf{1}_{\{\eta \leq t\}}, \mathbf{1}_{\{\delta \leq s\}} : s \leq t) \vee \mathcal{N}.
\end{aligned}$$

The information $\mathcal{G}^1$ corresponds to the case where upon retirement or death the insurer discards the previous health records of the insured. The sub-information $\mathcal{G}^2 \subset \mathcal{G}^1$ even keeps no record on the time of retirement. For most if not all practical purposes, the discarding of previous information upon death is of no importance.

Further, for describing the admissible information immediately before time $t \geq 0$ we define sequences of σ-algebras $\mathcal{G}^1_- = (\mathcal{G}^1_{t-})_{t\geq 0}$ and $\mathcal{G}^2_- = (\mathcal{G}^2_{t-})_{t\geq 0}$ by

$$\begin{aligned}
\mathcal{G}^1_{t-} &= \sigma(Z_s \mathbf{1}_{\{Z_{t-} \in \{1,\ldots,J\}\}}, \mathbf{1}_{\{\eta < s\}}, \mathbf{1}_{\{\delta < s\}} : s \leq t) \vee \mathcal{N}, \\
\mathcal{G}^2_{t-} &= \sigma(Z_s \mathbf{1}_{\{Z_{t-} \in \{1,\ldots,J\}\}}, \mathbf{1}_{\{\eta < t\}}, \mathbf{1}_{\{\delta < s\}} : s \leq t) \vee \mathcal{N}.
\end{aligned}$$

Lemma 2.2.4. *The σ-algebras $\mathcal{G}^1_t$, $\mathcal{G}^1_{t-}$, $\mathcal{G}^2_t$, and $\mathcal{G}^2_{t-}$, $t \geq 0$, can be brought on the form of* (2.2.1)–(2.2.2).

Proof. See Appendix 2.A. □

Lemma 2.2.4 gives a link to the general setting; note that the condition (2.2.3) is satisfied. From this point onward, for $\mathcal{G}^1$ and $\mathcal{G}^2$ the respective extended marked point process $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}}$ is always taken to be that from the proof of Lemma 2.2.4, see also Example 2.2.5 below.

Example 2.2.5. Let

$$\begin{aligned}
T_1 &= \eta, \quad & S_1 &= \infty, \quad & \zeta_1 &= (T_1, Z_{T_1}), \\
T_2 &= \delta, \quad & S_2 &= \infty, \quad & \zeta_2 &= (T_2, Z_{T_2}), \\
& & S_{2+i} &= T_1 \wedge T_2, \quad & \zeta_{2+i} &= (T_{2+i}, Z_{T_{2+i}}), \qquad i \in \mathbb{N},
\end{aligned}$$

and let T_{2+i}, $i \in \mathbb{N}$, be the jump times of the process counting the number of jumps of Z except retirement and death. Then according to the proof of Lemma 2.2.4, cf. Appendix 2.A,

$$\begin{aligned}\mathcal{G}^1_t &= \sigma(\{T_i \leq t < S_i\} \cap \{\zeta_i \in A\} : i \in \mathbb{N}, A \in \mathcal{E}) \vee \mathcal{N},\\ \mathcal{G}^1_{t-} &= \sigma(\{T_i < t \leq S_i\} \cap \{\zeta_i \in A\} : i \in \mathbb{N}, A \in \mathcal{E}) \vee \mathcal{N},\end{aligned}$$

for $t \geq 0$. The jump times have been permuted so that retirement and death have indices one and two, respectively. Consequently, the index of the jump time corresponding to retirement does not carry information concerning the total number of previous jumps. ∘

In the following, we develop a mathematically sound concept of state-wise prospective reserves in the case of non-monotone information, and we derive so-called stochastic Thiele equations describing the dynamics of state-wise prospective reserves in the presence of non-monotone information. The results are exemplified with non-monotone information given by $\mathcal{G}^1$ and $\mathcal{G}^2$, respectively, allowing us to discuss current actuarial practice regarding valuation of insurance liabilities in the presence of (possibly stochastic) retirement.

2.3 Prospective reserves in the presence of non-monotone information

In the case of monotone information, prospective reserves are so-called optional projections of accumulated future payments, suitably discounted. To our knowledge, there appears to be no unifying definition of general state-wise prospective reserves in the actuarial literature; in Norberg (1992), state-wise prospective reserves are given implicitly as prospective reserves evaluated on the relevant event, while Norberg (1996) in principle casts them based on somewhat arbitrary functional representations of prospective reserves. The properties of the state-wise prospective reserves as stochastic processes, including the existence and uniqueness of suitably regular versions, are not investigated. Furthermore, it is unclear from these proposals how to define state-wise prospective reserves in the presence of non-monotone information.

In this section, we present a sound and fruitful definition of state-wise prospective reserves in the presence of monotone as well as non-monotone information. In the presence of non-monotone information, the main idea is to take as underlying state process not Z giving the state of the insured but rather $\mathcal{Z}$ giving the state of information. The section is structured as follows. In Subsection 2.3.1, we introduce so-called state-wise counterparts and reveal the non-triviality of developing the concept of state-wise prospective reserves. In Subsection 2.3.2, we follow Christiansen (2020) on optional projections in the presence of non-monotone information, which turns out to be a fruitful Ansatz for a mathematically sound definition of state-wise quantities. Definitions of state-wise prospective reserves are introduced and discussed in Subsection 2.3.3.

2.3.1 State-wise counterparts

Suppose that $\mathcal{C} = (\mathcal{C}_t)_{t\geq 0}$ is a sequence of σ-algebras such that

$$\sigma(Z_t) \vee \mathcal{N} \subseteq \mathcal{C}_t \subseteq \mathcal{F}_t, \qquad t \geq 0.$$

Examples include $\mathcal{C} = \mathcal{G}$. We define sequences of families of sets $\mathcal{C}_j = (\mathcal{C}_{t,j})_{t\geq 0}$, $j \in S$, by

$$\mathcal{C}_{t,j} = \{A \in \mathcal{F}_{t-} : A \cap \{Z_t = j\} \in \mathcal{C}_t\}.$$

Lemma 2.3.1. *For each $(t, j) \in [0, \infty) \times \mathcal{Z}$ the family of sets $\mathcal{C}_{t,j}$ is a σ-algebra. Moreover,*

$$\mathcal{C}_t = \sigma(A \cap \{Z_t = j\} : A \in \mathcal{C}_{t,j}, j \in S)$$

for any $t \geq 0$.

Proof. Follows by standard set-theoretic calculations. $\square$

Example 2.3.2. Consider monotone information $\mathcal{F}$. Then $\mathcal{F}_{t,j} = \mathcal{F}_{t-}$ since $\mathcal{F}_{t-} \vee \sigma(Z_t) \subseteq \mathcal{F}_t$. $\circ$

Example 2.3.3. Consider the setting of Subsection 2.2.3. By defining

$$\begin{aligned}
\psi_j^1(s) &:= Z_s \mathbf{1}_{\{1,\ldots,J\}}(j) + \mathbf{1}_{\{\eta\leq s\}}\mathbf{1}_{\{J+1\}}(j) + (\mathbf{1}_{\{\eta\leq s\}}, \mathbf{1}_{\{\delta\leq s\}})\mathbf{1}_{\{J+2\}}(j), \\
\psi_j^2(s) &:= Z_s \mathbf{1}_{\{1,\ldots,J\}}(j) + \mathbf{1}_{\{\delta\leq s\}}\mathbf{1}_{\{J+2\}}(j)
\end{aligned}$$

for $s \geq 0$ and $j \in S$ we find that

$$\begin{aligned}
\mathcal{G}_{t,j}^1 &= \sigma(\psi_j^1(s) : s < t) \vee \mathcal{N}, \\
\mathcal{G}_{t,j}^2 &= \sigma(\psi_j^2(s) : s < t) \vee \mathcal{N}.
\end{aligned}$$

$\circ$

Let $Y = (Y(t))_{t\geq 0}$ be a real-valued stochastic process, and suppose that $Y(t)$ is $\mathcal{C}_t$-measurable for each $t \geq 0$. We now define the state-wise counterparts as follows:

Definition 2.3.4. *A family of real-valued stochastic processes $(Y_j)_{j\in S} = (Y_j(t))_{t\geq 0, j\in S}$ is said to be **state-wise counterparts** to Y if for each $(t, j) \in [0, \infty) \times S$:*

- *$Y_j(t)$ is $\mathcal{C}_{t,j}$-measurable,*
- *$\mathbf{1}_{\{Z_t=j\}} Y_j(t) \overset{a.s.}{=} \mathbf{1}_{\{Z_t=j\}} Y(t)$.*

In general, we suppress the dependency of state-wise counterparts on the specific sequence of σ-algebras $\mathcal{C}$.

Suppose for the moment that $Y^{\mathcal{G}} = (Y^{\mathcal{G}}(t))_{t\geq 0}$ is the prospective reserve under information $\mathcal{G}$ (to be formally defined later on). Then it is intuitively appealing to base the definition of the state-wise prospective reserves on the state-wise counterparts $(Y_j^{\mathcal{G}})_{j\in S}$ to $Y^{\mathcal{G}}$: they satisfy the key identity $\mathbf{1}_{\{Z_t=j\}}Y_j^{\mathcal{G}}(t) \overset{\text{a.s.}}{=} \mathbf{1}_{\{Z_t=j\}}Y^{\mathcal{G}}(t)$ and only rely on the information $\mathcal{G}_{t,j}$, which is the information available at time $t-$ that remains available at time t if $Z_t = j$.

For each $t \geq 0$ let m_t be the sub-probability measure that is uniquely defined on $\sigma(A \times \{j\} : A \in \mathcal{C}_{t,j}, j \in S)$ by

$$m_t(A \times \{j\}) = m_{t,j}(A) := P(A \cap \{Z_t = j\}), \qquad A \in \mathcal{C}_{t,j}, j \in S.$$

Proposition 2.3.5. *Let $Y = (Y(t))_{t\geq 0}$ be a real-valued stochastic process such that $Y(t)$ is integrable and $\mathcal{C}_t$-measurable for each $t \geq 0$. Then the state-wise counterparts $(Y_j)_{j\in S}$ to Y exist and for each $t \geq 0$ the mapping $\Omega \times S \ni (\omega, j) \mapsto Y_j(t)(\omega)$ is m_t-almost everywhere unique.*

Proof. See Appendix 2.A. □

The uniqueness of the state-wise counterparts does not extend beyond m_t-almost everywhere for fixed $t \geq 0$. In other words, viewed as processes the state-wise counterparts are not almost surely unique and thus not well-defined. Consequently, the definition of state-wise counterparts is mathematically flawed and it might therefore be unfortunate to base the definition of state-wise prospective reserves thereon.

Before we turn the attention to an alternative foundation based on an explicit representation of optional projections, we first present some results for the state-wise counterparts that are useful later.

Define a class of functionals $L_1(\Omega, \mathcal{A}, P) \ni X \mapsto \mathrm{E}_{t,j}[X \mid \mathcal{C}_{t,j}]$ by

$$\mathrm{E}_{t,j}[X \mid \mathcal{C}_{t,j}] := \frac{\mathrm{E}\big[X\mathbf{1}_{\{Z_t=j\}} \mid \mathcal{C}_{t,j}\big]}{\mathrm{E}\big[\mathbf{1}_{\{Z_t=j\}} \mid \mathcal{C}_{t,j}\big]},$$

where we impose the convention $0/0 := 0$. If $P(Z_t = j) > 0$, it holds that $\mathrm{E}_{t,j}[X \mid \mathcal{C}_{t,j}]$ are versions of the conditional expectations of $Y(t)$ given $\mathcal{C}_{t,j}$ w.r.t. the probability measure $P_{t,j}$ given by

$$P_{t,j}(A) = \frac{P(A \cap \{Z_t = j\})}{P(Z_t = j)}, \qquad A \in \mathcal{A},$$

cf. Exercise 34.4(a) of Billingsley (1994).

Based on similar techniques as in the proof of Proposition 2.3.5, one can then show that

$$Y_j(t) \overset{\text{a.s.}}{=} \mathrm{E}_{t,j}[Y(t) \mid \mathcal{C}_{t,j}] \,. \tag{2.3.1}$$

This provides an explicit representation of the state-wise counterparts.

We are now ready to derive the following law of iterated expectations:

Lemma 2.3.6. *Let $X \in L_1(\Omega, \mathcal{A}, P)$. Then for each $(t,j) \in [0,\infty) \times S$:*

$$\mathrm{E}_{t,j}[\mathrm{E}[X \mid \mathcal{C}_t] \mid \mathcal{C}_{t,j}] \overset{a.s.}{=} \mathrm{E}_{t,j}[X \mid \mathcal{C}_{t,j}] \,.$$

Proof. See Appendix 2.A. □

When $\mathcal{C}_{t,j}$ is generated by $F_{t,j} = f_{t,j}((Z_s)_{0 \leq s < t})$ added null sets $\mathcal{N}$ with $f_{t,j}$ some measurable function, it can be shown that

$$\mathrm{E}_{t,j}[Y(t) \mid \mathcal{C}_{t,j}] \overset{\text{a.s.}}{=} \mathrm{E}[Y(t) \mid F_{t,j}, Z_t = j] \,, \tag{2.3.2}$$

where the latter refers to path-wise integration w.r.t. the conditional distribution of $Y(t)$ given $(F_{t,j}, Z_t)$ and, further, evaluated in $\{F_{t,j}(\omega), j\}$. This provides an alternative explicit representation of the state-wise counterparts. Rewrites in the spirit of (2.3.2) are typical and occur frequently and opaquely in the remainder of the paper.

2.3.2 Optional projections and state-wise quantities

Let $Y = (Y(t))_{t \geq 0}$ be a real-valued stochastic process such that $Y(t)$ is integrable for each $t \geq 0$. If there exists an almost surely unique process $X = (X(t))_{t \geq 0}$ such that for each $t \geq 0$,

$$X(t) = \mathrm{E}\left[Y(t) \mid \mathcal{G}_t\right]$$

almost surely, then we denote $Y^{\mathcal{G}} := X$ as the optional projection of Y with respect to $\mathcal{G}$.

In the following we calculate conditional expectations given $(\boldsymbol{\zeta}_x, T_{xy}, \boldsymbol{\zeta}_{xy})$, $x, y \in \mathcal{S}, x \neq y$. We throughout assume that they are defined as path-wise integrals with respect to arbitrary but fixed regular conditional distributions $P(\cdot \mid \boldsymbol{\zeta}_x, T_{xy}, \boldsymbol{\zeta}_{xy})$. For a càdlàg or càglàd process $Y = (Y(t))_{\geq 0}$ with finite expected variation on compacts, let

$$
\begin{aligned}
\mathcal{Y}_x^{\mathcal{G}}(t) &:= \frac{\mathrm{E}[I_x(t)Y(t) \,|\, \boldsymbol{\zeta}_x]}{\mathrm{E}[I_x(t) \,|\, \boldsymbol{\zeta}_x]}, \quad t \geq 0, \\
\mathcal{Y}_x^{\mathcal{G}-}(t) &:= \frac{\mathrm{E}[I_x(t-)Y(t) \,|\, \boldsymbol{\zeta}_x]}{\mathrm{E}[I_x(t-) \,|\, \boldsymbol{\zeta}_x]}, \quad t > 0, \\
\mathcal{Y}_{xx}^{\mathcal{G}}(t) &:= \frac{\mathrm{E}[I_x(t-)I_x(t)Y(t) \,|\, \boldsymbol{\zeta}_x]}{\mathrm{E}[I_x(t-)I_x(t) \,|\, \boldsymbol{\zeta}_x]}, \quad t > 0, \\
\mathcal{Y}_{xy}^{\mathcal{G}-}(t,e) &:= \frac{\mathrm{E}[I_x(t-)Y(t) \,|\, \boldsymbol{\zeta}_x, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e]}{\mathrm{E}[I_x(t-) \,|\, \boldsymbol{\zeta}_x, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e]}, \quad x \neq y,\, e \in E_{xy},\, t > 0, \\
\mathcal{Y}_{xy}^{\mathcal{G}}(t,e) &:= \frac{\mathrm{E}[I_y(t)Y(t) \,|\, \boldsymbol{\zeta}_y, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e]}{\mathrm{E}[I_y(t) \,|\, \boldsymbol{\zeta}_y, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e]}, \quad x \neq y,\, e \in E_{xy},\, t > 0,
\end{aligned}
\tag{2.3.3}
$$

which are almost surely unique processes, cf. the discussion between Theorem 4.2 and Proposition 4.3 of Christiansen (2020). The above state-wise quantities refer to the state of information and changes in the state of information, rather than the state of the insured. In Subsection 2.3.3 we interpret these state-wise quantities when Y describes the accumulated future payments. The following proposition helps us in this regard.

Proposition 2.3.7. *Let Y be a càdlàg or càglàd process with finite expected variation on compacts. For each $t > 0$ we almost surely have*

$$
\begin{aligned}
I_x(t)\mathcal{Y}_x^{\mathcal{G}}(t) &= I_x(t)\,\mathrm{E}[Y(t) \,|\, \mathcal{G}_t], \\
I_x(t-)\mathcal{Y}_x^{\mathcal{G}-}(t) &= I_x(t-)\,\mathrm{E}[Y(t) \,|\, \mathcal{G}_{t-}], \\
I_x(t-)\mathcal{Y}_{xx}^{\mathcal{G}}(t) &= I_x(t-)\,\mathrm{E}[Y(t) \,|\, \mathcal{G}_{t-}, \mathcal{Z}_t = x], \\
I_x(t)\mathcal{Y}_{xx}^{\mathcal{G}}(t) &= I_x(t)\,\mathrm{E}[Y(t) \,|\, \mathcal{G}_t, \mathcal{Z}_{t-} = x], \\
I_x(t-)\mathcal{Y}_{xy}^{\mathcal{G}-}(t,e) &= I_x(t-)\ \mathrm{E}[Y(t) \,|\, \mathcal{G}_{t-}, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e], \\
I_y(t)\mathcal{Y}_{xy}^{\mathcal{G}}(t,e) &= I_y(t)\ \mathrm{E}[Y(t) \,|\, \mathcal{G}_t, T_{xy} = t, \boldsymbol{\zeta}_{xy} = e].
\end{aligned}
$$

Proof. See Proposition 4.3 and the proof of Theorem 4.2 in Christiansen (2020). □

The state-wise quantities $\mathcal{Y}_x^{\mathcal{G}}$ allow for a rather explicit characterization of the optional projection $Y^{\mathcal{G}}$:

Proposition 2.3.8. *Let $Y = (Y(t))_{t \geq 0}$ be a càdlàg process with with finite expected variation on compacts. Then the optional projection $Y^{\mathcal{G}}$ of Y exists and has the almost surely unique representation*

$$
Y^{\mathcal{G}}(t) = \sum_{x \in \mathcal{S}} I_x(t)\mathcal{Y}_x^{\mathcal{G}}(t), \quad t \geq 0.
$$

For each $x \in \mathcal{S}$ the processes $[0,\infty) \ni t \mapsto I_x(t)\mathcal{Y}_x^{\mathcal{G}}(t)$ and $(0,\infty) \ni t \mapsto I_x(t-)\mathcal{Y}_x^{\mathcal{G}}(t)$ have càdlàg modifications with paths of finite variation on compacts.

Proof. See Section 4 in Christiansen (2020). $\square$

In the special case of monotone information, we now establish a more direct relation between the different concepts of state-wise quantities. Setting $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}} := (\tau_i, \infty, (Z_{\tau_i}, \tau_i))_{i\in\mathbb{N}}$ we recover the filtration $\mathcal{F} = \mathcal{G}$. In this case, let

$$\begin{aligned} Y_{jk}^{\mathcal{F}-}(t) &:= \mathbf{1}_{\{Z_{t-}=j\}} \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} I_x(t-)\mathcal{Y}_{xy}^{\mathcal{F}-}(t,(k,t)), \\ Y_{jj}^{\mathcal{F}-}(t) &:= \mathbf{1}_{\{Z_{t-}=j\}} \sum_{x\in\mathcal{S}} I_x(t-)\mathcal{Y}_{xx}^{\mathcal{F}}(t) \end{aligned} \tag{2.3.4}$$

for $j,k \in S$, $j \neq k$, and $t > 0$.

Remark 2.3.9. In case of $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}} := (\tau_i, \infty, (Z_{\tau_i}, \tau_i))_{i\in\mathbb{N}}$, only those indicator processes I_x are different from constantly zero that have an x of the form $x = \{1,\dots,n\} \in \mathcal{S}$ for some $n \in \mathbb{N}_0$; here we define $\{1,\dots,n\}$ as the empty set in case of $n = 0$. In particular, we have

$$I_x(t-) = \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \quad \text{if } x = \{1,\dots,n\}$$

for $t > 0$ and with $\tau_0 := 0$. Moreover, the stopping times T_{xy} are only then different from constantly infinity if x and y are of the form $x = \{1,\dots,n\}$, $y = \{1,\dots,n+1\}$, $n \in \mathbb{N}_0$. In particular, for each $t > 0$ we almost surely have

$$\begin{aligned} &Y_{jk}^{\mathcal{F}-}(t) \\ &= \mathbf{1}_{\{Z_{t-}=j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \mathrm{E}[Y(t) \,|\, (Z_{\tau_1},\tau_1),\dots,(Z_{\tau_{n+1}},\tau_{n+1}) = (k,t)], \\ &Y_{jj}^{\mathcal{F}-}(t) \\ &= \mathbf{1}_{\{Z_{t-}=j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \frac{\mathrm{E}[\mathbf{1}_{\{Z_t=j\}} Y(t) \,|\, (Z_{\tau_1},\tau_1),\dots,(Z_{\tau_n},\tau_n)]}{\mathrm{E}[\mathbf{1}_{\{Z_t=j\}} \,|\, (Z_{\tau_1},\tau_1),\dots,(Z_{\tau_n},\tau_n)]} \end{aligned} \tag{2.3.5}$$

for $j,k \in S$, $j \neq k$. $\triangledown$

In the presence of monotone information, the following result relates the state-wise counterparts to the state-wise quantities introduced by (2.3.3).

Proposition 2.3.10. *Let $Y = (Y(t))_{t\geq 0}$ be a càdlàg process with finite expected variation on compacts. Denote by $Y^{\mathcal{F}}$ the corresponding optional projection and by $(Y_j^{\mathcal{F}})_{j\in S}$ the state-wise counterparts to $Y^{\mathcal{F}}$. At each time $t > 0$ it almost surely holds that*

$$Y_j^{\mathcal{F}}(t) = Y_{jj}^{\mathcal{F}-}(t) + \sum_{\substack{k\in S\\ j\neq k}} Y_{kj}^{\mathcal{F}-}(t)$$

for $j \in S$, where $Y_{jj}^{\mathcal{F}-}$ and $Y_{kj}^{\mathcal{F}-}$, $k \neq j$ are almost surely unique predictable processes defined by (2.3.4).

Proof. See Appendix 2.A. □

In the following, the notation $Y_j^{\mathcal{F}}$ always refers to the modification given by Proposition 2.3.10. Insisting on this essentially solves the issue of well-definedness of the state-wise counterparts in the presence of monotone information. In the general case, where we allow for non-monotone information, the issue persists. The next proposition contains results pertaining to the path properties of the modifications given by Proposition 2.3.10. The results ensure all later applications of e.g. integration by parts to be feasible.

Proposition 2.3.11. *For each $j \in S$ and almost each $\omega \in \Omega$ the path $t \mapsto Y_j^{\mathcal{F}}(t,\omega)$ is càdlàg and of finite variation on $[0,r] \cap [\tau_n(\omega), \tau_{n+1}(\omega)]$, $r > 0$, whenever $Z_{\tau_n}(\omega) = j$, $n \in \mathbb{N}_0$.*

Proof. See Appendix 2.A. □

Example 2.3.12. Consider the accumulated payments B, which is an $\mathcal{F}$-adapted càdlàg process with finite expected variation on compacts; in particular, $B^{\mathcal{F}} = B$. Proposition 2.3.10 yields

$$\begin{aligned} B(t) &= \sum_{j \in S} \mathbf{1}_{\{Z_t = j\}} \, B_j^{\mathcal{F}}(t) \\ &= \sum_{j \in S} \mathbf{1}_{\{Z_t = j\}} B_{jj}^{\mathcal{F}-}(t) + \sum_{\substack{j,k \in S \\ j \neq k}} \mathbf{1}_{\{Z_t = j\}} B_{kj}^{\mathcal{F}-}(t) \end{aligned}$$

almost surely for all $t > 0$. Recall that $B_{jj}^{\mathcal{F}-}(t) = \mathbf{1}_{\{Z_{t-} = j\}} B_{jj}^{\mathcal{F}-}(t)$ and $B_{jk}^{\mathcal{F}-}(t) = \mathbf{1}_{\{Z_{t-} = j\}} B_{jk}^{\mathcal{F}-}(t)$ for all $j,k \in S$, $j \neq k$. By applying integration by parts and rearranging the terms, one then finds

$$B(\mathrm{d}t) = \sum_{j \in S} \mathbf{1}_{\{Z_{t-} = j\}} B_{jj}^{\mathcal{F}-}(\mathrm{d}t) + \sum_{\substack{j,k \in S \\ j \neq k}} \left(B_{jk}^{\mathcal{F}-}(t) - B_{jj}^{\mathcal{F}-}(t)\right) N_{jk}(\mathrm{d}t) \tag{2.3.6}$$

almost surely. This recovers the classic decomposition into sojourn payments and transition payments in the following sense. Suppose the accumulated payments B are defined as

$$B(\mathrm{d}t) = \sum_{j \in S} \mathbf{1}_{\{Z_{t-} = j\}} \mathsf{B}_j(\mathrm{d}t) + \sum_{\substack{j,k \in S \\ j \neq k}} b_{jk}(t) \, N_{jk}(\mathrm{d}t),$$

where the cumulative sojourn payments B_j shall be $\mathcal{F}$-predictable càdlàg processes with finite expected variation on compacts and the transition payments b_{jk} shall be bounded and $\mathcal{F}$-predictable processes. By calculating $B_{jj}^{\mathcal{F}-}$ and $B_{jk}^{\mathcal{F}-}$ in (2.3.6) explicitly and comparing the results with the definition of B, one can show that

$$\mathbf{1}_{\{Z_{t-} = j\}} \mathsf{B}_j(\mathrm{d}t) = \mathbf{1}_{\{Z_{t-} = j\}} B_{jj}^{\mathcal{F}-}(\mathrm{d}t)$$

almost surely for $j \in S$ and for each $t > 0$,

$$\mathbf{1}_{\{Z_{t-}=j\}} b_{jk}(t) = \mathbf{1}_{\{Z_{t-}=j\}} \big(B_{jk}^{\mathcal{F}-}(t) - B_{jj}^{\mathcal{F}-}(t)\big)$$

almost surely for $j, k \in S$, $j \neq k$. By defining the process $\beta = (\beta(t))_{t \geq 0}$ via

$$\beta(t) = \sum_{\substack{j,k \in S \\ j \neq k}} b_{jk}(t)\, \Delta N_{jk}(t), \quad t \geq 0,$$

which equals the difference of a càdlàg and a càglàd process, we can alternatively recover the transition payments via the representation

$$\mathbf{1}_{\{Z_{t-}=j\}} b_{jk}(t) = \mathbf{1}_{\{Z_{t-}=j\}} \beta_{jk}^{\mathcal{F}-}(t),$$

which holds almost surely for all $t > 0$ and $j, k \in S$, $j \neq k$. ◦

2.3.3 State-wise prospective reserves

In the previous two subsections, we have introduced a range of state-wise concepts and quantities, including the state-wise counterparts, and we have studied their interrelation – in particular in the presence of monotone information. Building on this, we now turn our attention to mathematical sound definitions of state-wise prospective reserves. In the presence of monotone information, the definition bases on the concept of state-wise counterparts and refers to the state of the insured, while in the presence of non-monotone information, we rely on the state-wise quantities appearing in the explicit characterization of optional projections; these quantities refer to the state of information rather than the state of the insured.

Consider a deterministic bank account $\kappa : [0, \infty) \mapsto (0, \infty)$ assumed measurable, càdlàg, and of finite variation on compacts, with initial value $\kappa(0) = 1$. Denote with v the corresponding discount function given by

$$[0, \infty) \ni t \mapsto v(t) = \frac{1}{\kappa(t)}.$$

Denote from this point on by $Y = (Y(t))_{t \geq 0}$ the accumulated future payments, suitably discounted, given by

$$Y(t) = \int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s).$$

Note that Y has càdlàg sample paths of finite variation on compacts. We further suppose that $Y(t)$ has finite expected variation on compacts. This is for example the case if κ is bounded away from zero.

The prospective reserve under possibly non-monotone information is the almost surely unique optional projection $Y^{\mathcal{G}} = (Y^{\mathcal{G}}(t))_{t \geq 0}$ of Y w.r.t. $\mathcal{G}$ satisfying for each

$t \geq 0$

$$Y^{\mathcal{G}}(t) = \mathrm{E}\left[Y(t) \,|\, \mathcal{G}_t\right] = \mathrm{E}\left[\int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\middle|\, \mathcal{G}_t\right] \tag{2.3.7}$$

almost surely. This definition is consistent with the one proposed in Norberg (1991). State-wise prospective reserves are now defined as follows:

Definition 2.3.13. *For $j \in S$ the **classic state-wise prospective reserve in insured state j** is the not necessarily unique process $Y_j^{\mathcal{G}} = (Y_j^{\mathcal{G}}(t))_{t\geq 0}$, where $(Y_j^{\mathcal{G}})_{j\in S}$ are the state-wise counterparts to the prospective reserve $Y^{\mathcal{G}}$. For $x \in \mathcal{S}$, the **non-classic state-wise prospective reserve in information state x** is the almost surely unique process $\mathcal{Y}_x^{\mathcal{G}} = (\mathcal{Y}_x^{\mathcal{G}}(t))_{t\geq 0}$ given by*

$$\mathcal{Y}_x^{\mathcal{G}}(t) = \frac{\mathrm{E}[I_x(t)Y(t) \,|\, \boldsymbol{\zeta}_x]}{\mathrm{E}[I_x(t) \,|\, \boldsymbol{\zeta}_x]}$$

for $t \geq 0$.

In the following we shall follow the conventions of the literature and write $(V_j)_{j\in S}$ for the classic state-wise prospective reserves in the presence of monotone information $\mathcal{G} = \mathcal{F}$. Similarly, we write V for the prospective reserve in the presence of monotone information.

Note that for each $t \geq 0$, $j \in S$, and $x \in \mathcal{S}$, it almost surely holds that

$$\begin{aligned} \mathbf{1}_{\{Z_t=j\}} Y_j^{\mathcal{G}}(t) &= \mathbf{1}_{\{Z_t=j\}} Y^{\mathcal{G}}(t), \\ I_x(t)\mathcal{Y}_x^{\mathcal{G}}(t) &= I_x(t) Y^{\mathcal{G}}(t), \end{aligned}$$

cf. Definition 2.3.4 and Proposition 2.3.7. The proposed explicit definitions are therefore consistent with the implicit definition in the presence of monotone information put forward by Norberg (1992).

By an application of the law of iterated expectations, cf. Lemma 2.3.6, and the identity (2.3.1), we can for each $t \geq 0$ cast the classic state-wise prospective reserves as

$$Y_j^{\mathcal{G}}(t) \overset{\text{a.s.}}{=} \mathrm{E}\left[\int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\middle|\, \mathcal{G}_{t,j}, Z_t = j\right], \quad j \in S. \tag{2.3.8}$$

Example 2.3.14. Consider the case of monotone information $\mathcal{F}$, when by Example 2.3.2 we have $\mathcal{F}_{t,j} = \mathcal{F}_{t-}$. It follows that for each $t \geq 0$ and $j \in S$,

$$V_j(t) \overset{\text{a.s.}}{=} \mathrm{E}\left[\int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\middle|\, (Z_s)_{0\leq s<t}, Z_t = j\right],$$

cf. (2.3.8) and (2.3.2). ○

Example 2.3.15. Consider the framework of Subsection 2.2.3 with non-monotone information $\mathcal{G}^i$, $i \in \{1,2\}$, when by Example 2.3.3 we have $\mathcal{G}^i_{t,j} = \sigma(\psi^i_j(t)) \vee \mathcal{N}$. In the presence of non-monotone information $\mathcal{G}^i$, $i \in \{1,2\}$, we then for each $t \geq 0$ and $j \in S$ have

$$Y_j^{\mathcal{G}^i}(t) \overset{\text{a.s.}}{=} \mathrm{E}\left[\int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\middle|\, (\psi^i_j(s))_{0 \leq s < t}, Z_t = j\right],$$

cf. (2.3.8). For example,

$$Y_{J+1}^{\mathcal{G}^1}(t) \overset{\text{a.s.}}{=} \mathrm{E}\left[\int_{(t,\infty)} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\middle|\, U_t, Z_t = J+1\right],$$

where $U = (U_t)_{t\geq 0}$ is the duration process associated with Z.

Note that for each $t \geq 0$,

$$Y_j^{\mathcal{G}^i}(t) \overset{\text{a.s.}}{=} V_j(t)$$

for $j \in \{1,\dots,J\}$, while applying (2.3.2), (2.3.3), and the constructions of $\mathcal{G}^1$ and $\mathcal{G}^2$ according to the proof of Lemma 2.2.4, yields

$$\begin{aligned}
&Y_{J+1}^{\mathcal{G}^i}(t) \overset{\text{a.s.}}{=} \mathcal{Y}_{\{1\}}^{\mathcal{G}^i}(t), \\
&Y_{J+2}^{\mathcal{G}^i}(t) \overset{\text{a.s.}}{=} \mathbf{1}_{\{\eta \leq t\}} \mathcal{Y}_{\{1,2\}}^{\mathcal{G}^i}(t) + \mathbf{1}_{\{\eta > t\}} \mathcal{Y}_{\{2\}}^{\mathcal{G}^i}(t).
\end{aligned}$$

In the following, $(Y_j^{\mathcal{G}^i})_{j\in S}$ always refers to the modifications given by the above identities. Insisting on this ensures the classic state-wise prospective reserves to be well-defined in the presence of non-monotone information $\mathcal{G}^i$. ∘

As already discussed in Subsections 2.3.1–2.3.2, the state-wise counterparts are as a rule not well-defined as stochastic processes, since they are defined up to null-sets for an uncountable number of time points. In the presence of monotone information, $\mathcal{G} = \mathcal{F}$, we insist on taking the modification given by Proposition 2.3.10, which solves the problem of well-definedness, and in the following section we show how the concept of classic state-wise prospective reserves is sufficient to study dynamics of state-wise prospective reserves under monotone information. In the presence of non-monotone information, the classic state-wise prospective reserves are not well-defined as stochastic processes. Furthermore, as we show in the following section, the concept of non-classic state-wise prospective reserves, as well as the additional state-wise quantities given by (2.3.3), is necessary to study the dynamics of state-wise prospective reserves under non-monotone information. To develop the general theory of stochastic Thiele equations, we thus focus on the non-classic state-wise prospective reserves, which refer to the state of information. Still, when meaningful and relevant for specific instances of information, cf. Example 2.3.15, we cast the results in terms of the more intuitively appealing classic state-wise prospective reserves, which refer to the state of the insured.

In addition to the classic and non-classic state-wise prospective reserves, the additional state-wise quantities given by (2.3.3) prove useful. Based on Proposition 2.3.8 and Proposition 2.3.7, for each $x, y \in \mathcal{S}$, $x \neq y$, we interpret the state-wise quantities $\mathcal{Y}^{\mathcal{G}}_{xx}$, $\mathcal{Y}^{\mathcal{G}}_{xy}$, and $\mathcal{Y}^{\mathcal{G}-}_{xy}$ as follows:

- $\mathcal{Y}^{\mathcal{G}}_{xx}(t)$ is the prospective reserve for staying in information state x at time t: if in information state x at time $t-$ or time t, what one would set aside in case no change in information state occurs at time t,
- $\mathcal{Y}^{\mathcal{G}}_{xy}(t, e)$ is the *backward* prospective reserve at transition from information state x to information state y with information change e: if in information state y at time t, what one would set aside in case a change from information state x occurred with change in information e at exactly time t,
- $\mathcal{Y}^{\mathcal{G}-}_{xy}(t, e)$ is the *forward* prospective reserve at transition from information state x to y with information change e: if in information state x at time $t-$, what one would set aside in case a change to information state y occurs with change in information e at exactly time t.

2.4 Dynamics of state-wise prospective reserves

In this section, we present the main results of the paper by deriving so-called stochastic Thiele equations describing the dynamics of state-wise prospective reserves in the presence of non-monotone information. In principle, our method is based on the *infinitesimal approach* introduced and developed by Christiansen (2020) and relies on the explicit infinitesimal martingale representation theorem (see Theorem 6.1 and Theorem 7.1 in Christiansen, 2020). In comparison, stochastic Thiele equations in the presence of monotone information are closely related to the classic martingale representation theorem, see e.g. Norberg (1992) and Christiansen and Djehiche (2020).

In Subsection 2.4.1, we present and derive so-called infinitesimal forward/backward compensators describing the systematic part of the development of the state of information and the payments. Generalized stochastic Thiele equations are derived and interpreted in Subsection 2.4.2. Finally, in Subsection 2.4.3 we impose the specific framework of Subsection 2.2.3 with non-monotone information related to information discarding upon and after stochastic retirement and derive stochastic Thiele equations and – in the presence of certain intertemporal dependency structures – Feynman-Kac formulas exemplifying our results.

In the remainder of the paper, we generally suppose that

$$B(\mathrm{d}t) = \sum_{j\in S} \mathbf{1}_{\{Z_{t-}=j\}} b_j(t)\, \mu(\mathrm{d}t) + \sum_{\substack{j,k\in S\\ j\neq k}} b_{jk}(t)\, N_{jk}(\mathrm{d}t),$$

where b_j and b_{jk} are $\mathcal{F}$-predictable bounded processes and the measure μ is a sum of the Lebesgue-measure m and a countable number of Dirac-measures $(\epsilon_{t_n})_{n\in\mathbb{N}}$:

$$\mu(A) = m(A) + \sum_{n=1}^{\infty} \epsilon_{t_n}(A), \quad A \in \mathfrak{B}([0,\infty)),$$

for deterministic time points $0 \leq t_1 < t_2 < ...$ that are increasing to infinity (i.e. there are at most a finite number of such time points on each compact interval).

2.4.1 Infinitesimal compensators

The so-called compensator λ_{xy} of the random counting measure ν_{xy} is the unique $\mathcal{F}$-predictable random measure such that the difference $[0,\infty) \ni t \mapsto \nu_{xy}([0,t] \times A) - \lambda_{xy}([0,t] \times A)$ is an $\mathcal{F}$-martingale for each $A \in \mathfrak{B}(E_{xy})$. In particular, we have

$$\lambda_{xy}((0,t] \times A) = \lim_{n\to\infty} \sum_{\mathfrak{T}_n} \mathrm{E}[\nu_{xy}((t_k, t_{k+1}] \times A) \,|\, \mathcal{F}_{t_k}] \tag{2.4.1}$$

almost surely for each $t > 0$, where $(\mathfrak{T}_n)_{n\in\mathbb{N}}$ is any increasing sequence (i.e. $\mathfrak{T}_n \subset \mathfrak{T}_{n+1}$ for all n) of partitions $0 = t_0 < \cdots < t_n = t$ of the interval $[0,t]$ such that $|\mathfrak{T}_n| := \max\{t_k - t_{k-1} : k = 1, \ldots, n\} \to 0$ for $n \to \infty$. Christiansen (2020) expands this property to the non-motonone information $\mathcal{G}$ and denotes the random measures $\gamma_{xy}^{\mathcal{G}-}$ and $\gamma_{xy}^{\mathcal{G}}$ defined by

$$\gamma_{xy}^{\mathcal{G}-}((0,t] \times A) = \lim_{n\to\infty} \sum_{\mathfrak{T}_n} \mathrm{E}[\nu_{xy}((t_k, t_{k+1}] \times A) \,|\, \mathcal{G}_{t_k}], \quad t > 0,\ A \in \mathfrak{B}(E_{xy}),$$
$$\gamma_{xy}^{\mathcal{G}}((0,t] \times A) = \lim_{n\to\infty} \sum_{\mathfrak{T}_n} \mathrm{E}[\nu_{xy}((t_k, t_{k+1}] \times A) \,|\, \mathcal{G}_{t_{k+1}}], \quad t > 0,\ A \in \mathfrak{B}(E_{xy}),$$

as infinitesimal forward compensator (IF-compensator) and infinitesimal backward compensator (IB-compensator) of ν_{xy} with respect to $\mathcal{G}$, given that the limits exist for all $t > 0$ almost surely.

In the special case of monotone information $\mathcal{G} = \mathcal{F}$ the IF-compensator equals the classic compensator and the IB-compensator equals the counting measure itself, i.e. $\gamma_{xy}^{\mathcal{F}-} = \lambda_{xy}$ and $\gamma_{xy}^{\mathcal{F}} = \nu_{xy}$ almost surely.

Proposition 2.4.1. *For each $x, y \in \mathcal{S}$, $x \neq y$, the IF-compensator $\gamma_{xy}^{\mathcal{G}-}$ and the IB-compensator $\gamma_{xy}^{\mathcal{G}}$ of ν_{xy} exist and satisfy*

$$\gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e) = I_x(t-) g_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e), \qquad \gamma_{xy}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) = I_y(t) g_{xy}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e),$$
$$g_{xy}^{\mathcal{G}-}((0,t] \times A) := \int_{(0,t] \times A} \frac{\mathbf{1}_{\{\mathrm{E}[I_x(s-)\,|\,\boldsymbol{\zeta}_x]>0\}}}{\mathrm{E}[I_x(s-)\,|\,\boldsymbol{\zeta}_x]} P((T_{xy}, \boldsymbol{\zeta}_{xy}) \in \mathrm{d}s \times \mathrm{d}e\,|\,\boldsymbol{\zeta}_x),$$
$$g_{xy}^{\mathcal{G}}((0,t] \times A) := \int_{(0,t] \times A} \frac{\mathbf{1}_{\{\mathrm{E}[I_y(s)\,|\,\boldsymbol{\zeta}_y]>0\}}}{\mathrm{E}[I_y(s)\,|\,\boldsymbol{\zeta}_y]} P((T_{xy}, \boldsymbol{\zeta}_{xy}) \in \mathrm{d}s \times \mathrm{d}e\,|\,\boldsymbol{\zeta}_y),$$

almost surely (with $A \in \mathfrak{B}(E_{xy})$, $t > 0$).

Proof. See Proposition 5.1 and Theorem 5.2 in Christiansen (2020). □

Denote by b the sojourn payment rate given by

$$b(t) = \sum_{j \in S} \mathbf{1}_{\{Z_{t-}=j\}} b_j(t), \quad t > 0,$$

and denote by β the transition payments given by

$$\beta(t) = \sum_{\substack{j,k \in S \\ j \neq k}} b_{jk}(t)\, \Delta N_{jk}(t), \quad t > 0.$$

Proposition 2.4.2. *The payment process B has an IF-compensator $C_B^{\mathcal{G}-}$ with respect to $\mathcal{G}$ of the form*

$$C_B^{\mathcal{G}-}(\mathrm{d}t) \overset{a.s.}{=} \sum_{x \in \mathcal{S}} I_x(t-) b_x^{\mathcal{G}-}(t)\, \mu(\mathrm{d}t) + \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} \int_{E_{xy}} \beta_{xy}^{\mathcal{G}-}(t,e)\, \gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e),$$

where $b_x^{\mathcal{G}-}$ and $\beta_{xy}^{\mathcal{G}-}$ are the processes defined from b and β by the second and fourth line in (2.3.3)*, respectively.*

Proof. See Theorem 5.2 and Example 7.2 in Christiansen (2020). Note that (2.2.4) holds and that β can be decomposed into a sum of a càdlàg and a càglàd process both with finite expected variation on compacts. □

Applying similar techniques as in the proof of Proposition 2.4.1 and the proof of Proposition 2.4.2, one can show that if for all $t > 0$ each $b_j(t)$ and $b_{jk}(t)$ is $\mathcal{G}_{t-}$-measurable, then

$$C_B^{\mathcal{G}-}(\mathrm{d}t) \overset{\text{a.s.}}{=} \sum_{j \in S} \mathbf{1}_{\{Z_{t-}=j\}} b_j(t)\, \mu(\mathrm{d}t) + \sum_{\substack{j,k \in S \\ j \neq k}} b_{jk}(t)\, \Gamma_{jk}^{\mathcal{G}-}(\mathrm{d}t),$$

where $\Gamma^{\mathcal{G}-}$ are the IF-compensators of the multivariate counting process N (associated with Z) w.r.t. $\mathcal{G}$.

In general, we thus interpret $b_x^{\mathcal{G}-}$ as the ($\mathcal{G}$-averaged) sojourn payments in information state $x \in \mathcal{S}$ and $\beta_{xy}^{\mathcal{G}-}(\cdot, e)$ as the ($\mathcal{G}$-averaged) transition payment for a change in information e from information state x to information state y.

2.4.2 Stochastic Thiele equations

We are now ready to present stochastic differential equations describing the dynamics of the non-classic state-wise prospective reserves $(\mathcal{Y}_x^{\mathcal{G}})_{x \in \mathcal{S}}$ in the presence of general non-monotone information $\mathcal{G}$:

Theorem 2.4.3 (Generalized stochastic Thiele equation)**.** *The non-classic state-wise prospective reserves $(\mathcal{Y}_x^{\mathcal{G}})_{x \in \mathcal{S}}$ almost surely satisfy the stochastic differential equation*

$$
\begin{aligned}
0 = \sum_{x \in \mathcal{S}} I_x(t-) \Bigg(\mathcal{Y}_x^{\mathcal{G}}(\mathrm{d}t) - \mathcal{Y}_x^{\mathcal{G}}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b_x^{\mathcal{G}-}(t)\, \mu(\mathrm{d}t) \\
+ \sum_{y : y \neq x} \int_{E_{xy}} \mathcal{R}^{\mathcal{G}-}(t, x, y, e)\, g_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e) \\
- \sum_{y : y \neq x} \int_{E_{yx}} \mathcal{R}^{\mathcal{G}}(t, y, x, e)\, g_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) \Bigg),
\end{aligned}
\tag{2.4.2}
$$

where for $x, y \in \mathcal{S}$, $x \neq y$,

$$
\begin{aligned}
\mathcal{R}^{\mathcal{G}-}(t, x, y, e) &:= \beta_{xy}^{\mathcal{G}-}(t, e) + \mathcal{Y}_{xy}^{\mathcal{G}-}(t, e) - \mathcal{Y}_{xx}^{\mathcal{G}}(t), \\
\mathcal{R}^{\mathcal{G}}(t, y, x, e) &:= \mathcal{Y}_{yx}^{\mathcal{G}}(t, e) - \mathcal{Y}_{xx}^{\mathcal{G}}(t).
\end{aligned}
$$

Remark 2.4.4. According to Proposition 2.4.1, we might replace $g_{xy}^{\mathcal{G}-}$ by $\gamma_{xy}^{\mathcal{G}-}$ in (2.4.2). In the following, we prefer this representation. Note that we are (in general) unable to replace $g_{yx}^{\mathcal{G}}$ by $\gamma_{yx}^{\mathcal{G}}$. $\triangledown$

In the presence of monotone information $\mathcal{G} = \mathcal{F}$, starting from Theorem 2.4.3 one can derive the following stochastic differential equations describing the dynamics of the classic state-wise prospective reserves $(V_j)_{j \in S}$:

Corollary 2.4.5 (Classic stochastic Thiele equation)**.** *The classic state-wise prospective reserves $(V_j)_{j \in S}$ almost surely satisfy the stochastic differential equation*

$$
\begin{aligned}
0 = \sum_{j \in S} \mathbf{1}_{\{Z_{t-} = j\}} \Bigg(V_j(\mathrm{d}t) - V_j(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b_j(t)\, \mu(\mathrm{d}t) \\
+ \sum_{k : k \neq j} R_{jk}(t)\, \Lambda_{jk}(\mathrm{d}t) \Bigg),
\end{aligned}
\tag{2.4.3}
$$

where $R_{jk}(t) := b_{jk}(t) + V_k(t) - V_j(t)$ are the classic sum at risks and where $\Lambda_{jk} := \Gamma_{jk}^{\mathcal{F}-}$ are the classic $\mathcal{F}$-compensators of the multivariate counting process N.

Before we present the proofs of Theorem 2.4.3 and Corollary 2.4.5, we first provide an interpretation of the results. In the presence of monotone information, Corollary 2.4.5 yields stochastic differential equations that are directly comparable to the stochastic Thiele equations of Norberg (1992, 1996). In Norberg (1992, 1996), the $\mathcal{F}$-compensators Λ of N are assumed to admit densities w.r.t. the Lebesgue-measure, and the result is derived by suitably applying the martingale representation theorem and identifying the integrands. The method of the present paper, while extended to also cover the non-monotone case, is based on a suitable application of the explicit infinitesimal martingale representation theorem. In particular, Corollary 2.4.5 can also be derived directly from the classic martingale representation theorem following Christiansen and Djehiche (2020); in this case, the restriction to slightly less general payments, cf. beginning of Section 2.4, is not necessary.

The stochastic differential equation of Theorem 2.4.3 is in a twofold manner fundamentally different from the stochastic Thiele equation in the presence of monotone information. Firstly, the sum at risks appearing in the term involving the IF-compensators, which correspond to ordinary compensators in the presence of monotone information, take a different form. Rather than being the difference of two state-wise prospective reserves added the relevant transition payment, it involves the difference of the forward state-wise prospective reserve and the prospective reserve for staying in the state added relevant transition payment. In the presence of monotone information, we can show that the forward state-wise prospective reserve and the prospective reserve for staying in the state can be replaced by relevant ordinary state-wise prospective reserves, but this is not necessarily the case in the presence of non-monotone information. Here the possibility of information discarding entails a possible improvement in the accuracy of the reserving by utilizing the information available at time $t-$ and time t, rather than utilizing only the information available at time t.

Secondly, the stochastic differential equation of Theorem 2.4.3 contains an additional term involving $g_{yx}^{\mathcal{G}}$, $y \neq x$, and thus relates to the IB-compensators. In the presence of monotone information, we can show that this term is zero. It is the backward looking equivalent of the term involving the IF-compensators. Based on the information currently available, the term adjusts the dynamics to take into account the possibility that information discarding has just occurred.

In Subsection 2.4.3, we derive and interpret stochastic Thiele equations in the presence of specific examples of non-monotone information related to stochastic retirement. We refer to this subsection for further interpretation and discussion of the general results.

Proof of Theorem 2.4.3. Analogously to Proposition 2.4.2, one can show that the discounted payment process $\bar{B}$ given by $\bar{B}(0) = B(0)$ and $\bar{B}(\mathrm{d}t) := v(t)B(\mathrm{d}t)$ admits the IF-compensator

$$C_{\bar{B}}^{\mathcal{G}-}(\mathrm{d}t) = \sum_{x \in \mathcal{S}} I_x(t-) v(t)\, b_x^{\mathcal{G}-}(t)\, \mu(\mathrm{d}t) + \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} \int_{E_{xy}} v(t)\, \beta_{xy}^{\mathcal{G}-}(t,e)\, \gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e).$$

According to Theorem 7.1 in Christiansen (2020), the process $[0,\infty) \ni t \mapsto \bar{Y}(t) = v(t)Y(t)$ almost surely satisfies the equation

$$\begin{aligned} \bar{Y}^{\mathcal{G}}(\mathrm{d}t) = -C_{\bar{B}}^{\mathcal{G}-}(\mathrm{d}t) &+ \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} \int_{E_{xy}} v(t)\big(\mathcal{Y}_{xy}^{\mathcal{G}-}(t,e) - \mathcal{Y}_{xx}^{\mathcal{G}}(t)\big)(\nu_{xy} - \gamma_{xy}^{\mathcal{G}-})(\mathrm{d}t \times \mathrm{d}e) \\ &- \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} \int_{E_{xy}} v(t)\big(\mathcal{Y}_{xy}^{\mathcal{G}}(t,e) - \mathcal{Y}_{yy}^{\mathcal{G}}(t)\big)(\nu_{xy} - \gamma_{xy}^{\mathcal{G}})(\mathrm{d}t \times \mathrm{d}e). \end{aligned}$$

On the other hand, by applying integration by parts on $\bar{Y}(t) \overset{\text{a.s.}}{=} \sum_{x \in \mathcal{S}} I_x(t) \bar{\mathcal{Y}}_x^{\mathcal{G}}(t)$ and using $\bar{Y}^{\mathcal{G}} \overset{\text{a.s.}}{=} v(t) Y^{\mathcal{G}}$, we can show that

$$\bar{Y}^{\mathcal{G}}(\mathrm{d}t) \overset{\text{a.s.}}{=} \sum_{x \in \mathcal{S}} I_x(t-) \bar{\mathcal{Y}}_x^{\mathcal{G}}(\mathrm{d}t) + \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} v(t)\big(\mathcal{Y}_y^{\mathcal{G}}(t) - \mathcal{Y}_x^{\mathcal{G}}(t)\big)\nu_{xy}(\mathrm{d}t \times E_{xy}).$$

Thus, by equating the latter two equations and rearranging the terms, while using the fact that $\gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e) = I_x(t-)\gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e)$ and $\gamma_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) = I_x(t)\gamma_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e)$ almost surely and the equation $I_x(t) = I_x(t-)I_x(t) + \mathbf{1}_{\{\mathcal{Z}_{t-} \neq x\}} I_x(t)$, we obtain

$$\begin{aligned} 0 \overset{\text{a.s.}}{=} \sum_{x \in \mathcal{S}} I_x(t-) \Bigg(& \bar{\mathcal{Y}}_x^{\mathcal{G}}(\mathrm{d}t) + v(t) b_x^{\mathcal{G}-}(t)\, \mu(\mathrm{d}t) \\ & + \sum_{y : y \neq x} \int_{E_{xy}} v(t) \mathcal{R}^{\mathcal{G}-}(t,x,y,e)\, \gamma_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e) \\ & - \sum_{y : y \neq x} \int_{E_{yx}} v(t) \mathcal{R}^{\mathcal{G}}(t,y,x,e)\, \gamma_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) \Bigg), \\ - \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} & v(t)\big(\mathcal{Y}_{xy}^{\mathcal{G}-}(t,e) - \mathcal{Y}_{xy}^{\mathcal{G}}(t,e)\big)\nu_{xy}(\mathrm{d}t \times E_{xy}) \\ + \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} & v(t)\big(\mathcal{Y}_{xx}^{\mathcal{G}}(t) - \mathcal{Y}_{yy}^{\mathcal{G}}(t)\big)\nu_{xy}(\mathrm{d}t \times E_{xy}) \\ - \sum_{\substack{x,y \in \mathcal{S} \\ y \neq x}} & \int_{E_{yx}} v(t)\big(\mathcal{Y}_{yx}^{\mathcal{G}}(t,e) - \mathcal{Y}_{xx}^{\mathcal{G}}(t)\big)\mathbf{1}_{\{\mathcal{Z}_{t-} \neq x\}} I_x(t)\, \gamma_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) \\ + \sum_{\substack{x,y \in \mathcal{S} \\ x \neq y}} & v(t)\big(\mathcal{Y}_y^{\mathcal{G}}(t) - \mathcal{Y}_x^{\mathcal{G}}(t)\big)\nu_{xy}(\mathrm{d}t \times E_{xy}). \end{aligned}$$

Proposition 2.4.1 and the identity

$$g_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) \overset{\text{a.s.}}{=} \mathbf{1}_{\{\mathcal{Z}_t \neq x\}} g_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e) + \gamma_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e)$$

then yield

$$
\begin{aligned}
0 \overset{\text{a.s.}}{=} \sum_{x\in\mathcal{S}} I_x(t-)\Bigg(&\bar{\mathcal{Y}}^{\mathcal{G}}_x(\mathrm{d}t) + v(t) b^{\mathcal{G}-}_x(t)\,\mu(\mathrm{d}t) \\
&+ \sum_{y:y\neq x} \int_{E_{xy}} v(t)\mathcal{R}^{\mathcal{G}-}(t,x,y,e)\, g^{\mathcal{G}-}_{xy}(\mathrm{d}t\times\mathrm{d}e) \\
&- \sum_{y:y\neq x} \int_{E_{yx}} v(t)\mathcal{R}^{\mathcal{G}}(t,y,x,e)\, g^{\mathcal{G}}_{yx}(\mathrm{d}t\times\mathrm{d}e)\Bigg), \\
&- \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)\big(\mathcal{Y}^{\mathcal{G}-}_{xy}(t,e) - \mathcal{Y}^{\mathcal{G}}_{xy}(t,e)\big)\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
&+ \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)\big(\mathcal{Y}^{\mathcal{G}}_{xx}(t) - \mathcal{Y}^{\mathcal{G}}_{yy}(t)\big)\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
&- \sum_{\substack{x,y\in\mathcal{S}\\ y\neq x}} \int_{E_{yx}} v(t)\big(\mathcal{Y}^{\mathcal{G}}_{yx}(t,e) - \mathcal{Y}^{\mathcal{G}}_{xx}(t)\big)\mathbf{1}_{\{\mathcal{Z}_{t-}\neq x\}} I_x(t)\, g^{\mathcal{G}}_{yx}(\mathrm{d}t\times\mathrm{d}e) \\
&+ \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)\big(\mathcal{Y}^{\mathcal{G}}_{y}(t) - \mathcal{Y}^{\mathcal{G}}_{x}(t)\big)\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
&+ \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} \int_{E_{yx}} v(t)\big(\mathcal{Y}^{\mathcal{G}}_{yx}(t,e) - \mathcal{Y}^{\mathcal{G}}_{xx}(t)\big) I_x(t-)\mathbf{1}_{\{\mathcal{Z}_{t}\neq x\}} g^{\mathcal{G}}_{yx}(\mathrm{d}t\times\mathrm{d}e).
\end{aligned}
$$

The fourth line equals zero because of (6.7) in Christiansen (2020). The fifth, sixth, seventh, and eighth line together equal

$$
\begin{aligned}
&\sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)(\mathcal{Y}^{\mathcal{G}}_{y}(t) - \mathcal{Y}^{\mathcal{G}}_{yy}(t))\,\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
&- \sum_{\substack{x,y\in\mathcal{S}\\ y\neq x}} \Bigg(\int_{E_{yx}} \bigg\{ v(t)\big(\mathcal{Y}^{\mathcal{G}}_{yx}(t,e) - \mathcal{Y}^{\mathcal{G}}_{xx}(t)\big) \\
&\qquad\qquad \sum_{z:z\neq x} \big(\nu_{zx}(\{t\}\times E_{zx}) - \nu_{xz}(\{t\}\times E_{xz})\big)\bigg\} g^{\mathcal{G}}_{yx}(\mathrm{d}t\times\mathrm{d}e)\Bigg) \\
&+ \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)(\mathcal{Y}^{\mathcal{G}}_{xx}(t) - \mathcal{Y}^{\mathcal{G}}_{x}(t))\,\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
= &\sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)(\mathcal{Y}^{\mathcal{G}}_{x}(t) - \mathcal{Y}^{\mathcal{G}}_{xx}(t))\,\nu_{yx}(\mathrm{d}t\times E_{yx}) + \sum_{\substack{x,y\in\mathcal{S}\\ x\neq y}} v(t)(\mathcal{Y}^{\mathcal{G}}_{xx}(t) - \mathcal{Y}^{\mathcal{G}}_{x}(t))\,\nu_{xy}(\mathrm{d}t\times E_{xy}) \\
&- \sum_{\substack{x,y\in\mathcal{S}\\ y\neq x}} \Bigg(\sum_{z:z\neq x} \int_{E_{zx}} v(t)\big(\mathcal{Y}^{\mathcal{G}}_{zx}(t,e) - \mathcal{Y}^{\mathcal{G}}_{xx}(t)\big)\, g^{\mathcal{G}}_{zx}(\{t\}\times\mathrm{d}e) \\
&\qquad \big(\nu_{yx}(\mathrm{d}t\times E_{yx}) - \nu_{xy}(\mathrm{d}t\times E_{xy})\big)\Bigg)
\end{aligned}
$$

almost surely, because $\sum_{z:z\neq x} \nu_{zx}(\{t\} \times E_{zx})$ and $\sum_{z:z\neq x} \nu_{xz}(\{t\} \times E_{xz})$ are almost surely non-zero only at finitely many time points. The latter three lines also add up to zero since

$$\mathcal{Y}_x^{\mathcal{G}}(t) = \mathcal{Y}_{xx}^{\mathcal{G}}(t)\Big(1 - \sum_{z:z\neq x} g_{zx}^{\mathcal{G}}(\{t\} \times E_{zx})\Big) + \sum_{z:z\neq x} \int_{E_{zx}} \mathcal{Y}_{zx}^{\mathcal{G}}(t,e)\, g_{zx}^{\mathcal{G}}(\{t\} \times \mathrm{d}e)$$

almost surely. This identity is a consequence of the following observations. If $\mathrm{E}[I_x(s) \,|\, \boldsymbol{\zeta}_x] = 0$, then by definition, $\mathcal{Y}_{xx}^{\mathcal{G}}(t) = 0$, $\mathcal{Y}_{xx}^{\mathcal{G}}(t) = 0$, and $g_{zx}^{\mathcal{G}}(\{t\} \times \mathrm{d}e) = 0$ almost surely and the identity simply reads $0 = 0$. On the other hand, if $\mathrm{E}[I_x(s) \,|\, \boldsymbol{\zeta}_x] > 0$, then by applying Proposition 2.3.7, (2.2.5), and Proposition 2.4.1,

$$\begin{aligned}
\mathcal{Y}_x^{\mathcal{G}}(t) &= \mathrm{E}[Y(t) \,|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x] \\
&= \mathrm{E}[Y(t) I_x(t-) \,|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x] + \mathrm{E}\left[Y(t) \sum_{z:z\neq x} I_z(t-) \,\middle|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x\right] \\
&= \mathrm{E}[Y(t) \,|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x, \mathcal{Z}_{t-} = x]\ \mathrm{E}[I_x(t-) \,|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x] \\
&\quad + \sum_{z:z\neq x} \int_{E_{zx}} \mathrm{E}[Y(t) \,|\, \boldsymbol{\zeta}_x, \mathcal{Z}_t = x, T_{zx} = t, \boldsymbol{\zeta}_{zx} = e]\, g_{zx}^{\mathcal{G}}(\{t\} \times \mathrm{d}e) \\
&= \mathcal{Y}_{xx}^{\mathcal{G}}(t)\Big(1 - \sum_{z:z\neq x} g_{zx}^{\mathcal{G}}(\{t\} \times E_{zx})\Big) + \sum_{z:z\neq x} \int_{E_{zx}} \mathcal{Y}_{zx}^{\mathcal{G}}(t,e)\, g_{zx}^{\mathcal{G}}(\{t\} \times \mathrm{d}e)
\end{aligned}$$

almost surely.

All in all, we have

$$\begin{aligned}
0 = \sum_{x\in\mathcal{S}} I_x(t-)\Bigg(& \bar{\mathcal{Y}}_x^{\mathcal{G}}(\mathrm{d}t) + v(t) b_x^{\mathcal{G}-}(t)\mu(\mathrm{d}t) \\
&+ \sum_{y:y\neq x} \int_{E_{xy}} v(t)\mathcal{R}^{\mathcal{G}-}(t,x,y,e)\, g_{xy}^{\mathcal{G}-}(\mathrm{d}t \times \mathrm{d}e) \\
&- \sum_{y:y\neq x} \int_{E_{xy}} v(t)\mathcal{R}^{\mathcal{G}}(t,x,y,e)\, g_{yx}^{\mathcal{G}}(\mathrm{d}t \times \mathrm{d}e)\Bigg).
\end{aligned}$$

Now apply integration by parts on $\bar{\mathcal{Y}}_x^{\mathcal{G}}(t) = v(t)\mathcal{Y}_x^{\mathcal{G}}(t)$ and rearrange the terms in order to end up with the statement of the theorem. □

Proof of Corollary 2.4.5. By setting $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}} = (\tau_i, \infty, (\tau_i, Z_{\tau_i}))_{i\in\mathbb{N}}$ we obtain $\mathcal{G} = \mathcal{F}$ such that $(\mathcal{Y}_x^{\mathcal{F}})_{x\in\mathcal{S}}$ satisfy (2.4.2) almost surely. Since $\gamma_{yx}^{\mathcal{F}} \overset{\text{a.s.}}{=} \nu_{yx}$, we must have $I_x(t-) g_{yx}^{\mathcal{F}}(\mathrm{d}t \times \mathrm{d}e) \overset{\text{a.s.}}{=} 0$ when

$$I_x(t-) \sum_{y:y\neq x} \int_{E_{yx}} \mathcal{R}^{\mathcal{F}}(t,x,y,e)\, g_{yx}^{\mathcal{F}}(\mathrm{d}t \times \mathrm{d}e) = 0$$

almost surely. By Remark 2.3.9 and starting from (2.4.2), similar arguments as in the proof of Proposition 2.3.10 yield the following stochastic differential equations:

$$\sum_{n=0}^{\infty} \mathbf{1}_{\{Z_{\tau_n}=j\}} \mathbf{1}_{\{\tau_n<t\leq\tau_{n+1}\}} \mathcal{Y}^{\mathcal{F}}_{\{1,\ldots,n\}}(\mathrm{d}t)$$
$$\overset{\text{a.s}}{=} \mathbf{1}_{\{Z_{t-}=j\}} \bigg(V_j(t-)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_j(t)\,\mu(\mathrm{d}t) - \sum_{k:k\neq j} \big(b_{jk}(t) + V_k(t) - V_j(t)\big)\,\Gamma^{\mathcal{F}-}_{jk}(\mathrm{d}t)\bigg)$$

for $j \in S$. By tedious yet straightforward calculations, it is possible to show that

$$\sum_{n=0}^{\infty} \big(V_j(t) - \mathcal{Y}^{\mathcal{F}}_{\{1,\ldots,n\}}\big)\,\mathrm{d}\big(\mathbf{1}_{\{Z_{\tau_n}=j\}}\mathbf{1}_{\{\tau_n\leq t<\tau_{n+1}\}}\big) \overset{\text{a.s.}}{=} 0, \quad j \in S,$$

which implies

$$\sum_{n=0}^{\infty} \mathbf{1}_{\{Z_{\tau_n}=j\}} \mathbf{1}_{\{\tau_n<t\leq\tau_{n+1}\}} \mathcal{Y}^{\mathcal{F}}_{\{1,\ldots,n\}}(\mathrm{d}t) \overset{\text{a.s.}}{=} \mathbf{1}_{\{Z_{t-}=j\}} V_j(\mathrm{d}t), \quad j \in S,$$

by an application of integration by parts. Collecting results establishes the desired result. □

In the case where the payments B themselves depend on the prospective reserve V, the (stochastic) Thiele equations rather than (2.3.7) might serve as definition for the prospective reserve V, see e.g. Djehiche and Löfdahl (2016) and Christiansen and Djehiche (2020). In the presence of monotone information, this point of view is encapsulated by the following result.

Proposition 2.4.6. *Let there be a maximal contract time $n < \infty$, i.e. each b_j and b_{jk} is constantly zero on the interval (n, ∞). Suppose that W_j, $j \in S$, are $\mathcal{F}$-predictable bounded processes such that $[0,\infty) \ni t \mapsto \mathbf{1}_{\{Z_t=j\}} W_j(t)$ almost surely has càdlàg paths for all $j \in S$. If W_j, $j \in S$, satisfy the stochastic differential equations*

$$\begin{aligned} 0 = \mathbf{1}_{\{Z_{t-}=j\}} \bigg(& W_j(\mathrm{d}t) - W_j(t-)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b_j(t)\,\mu(\mathrm{d}t) \\ & + \sum_{k:k\neq j} (b_{jk}(t) + W_k(t) - W_j(t))\,\Lambda_{jk}(\mathrm{d}t)\bigg) \end{aligned} \tag{2.4.4}$$

with terminal condition $W_j(n) = 0$, $j \in S$, then $W_{Z_t}(t) = V(t)$ almost surely for all $t \in [0, n]$.

Proof. Suppose that $[\tau_n, \tau_{n+1})$ is an interval where $Z_t = j$. Then W_j is càdlàg on $[\tau_n, \tau_{n+1}]$ because of our càdlàg assumption for $\mathbf{1}_{\{Z_t=j\}} W_j(t)$ and since the value of W_j at the right end point τ_{n+1} is not relevant for the càdlàg property. Furthermore,

W_j has paths of finite variation on $[\tau_n, \tau_{n+1}]$, since the stochastic differential equation implies the finite variation property on $(\tau_n, \tau_{n+1}]$ and since adding the left end point does not change the finite variation property. By applying integration by parts and the stochastic differential equations for the processes W_j, $j \in S$, we obtain

$$\begin{aligned}
\mathrm{d}\big(v(t)W_{Z_t}(t)\big) &= \sum_{j\in S} \mathbf{1}_{\{Z_{t-}=j\}} \bigg(v(t)W_j(\mathrm{d}t) - W_j(t-)v(t)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)}\bigg) \\
&\quad + \sum_{\substack{j,k\in S\\ j\neq k}} v(t)(W_k(t) - W_j(t))\, N_{jk}(\mathrm{d}t) \\
&= -\, v(t)\, B(\mathrm{d}t) + \sum_{\substack{j,k\in S\\ j\neq k}} v(t)(b_{jk}(t) + W_k(t) - W_j(t))(N_{jk} - \Lambda_{jk})(\mathrm{d}t)
\end{aligned}$$

almost surely. Since each $[0,\infty) \ni t \mapsto b_{jk}(t) + W_k(t) - W_j(t)$ is $\mathcal{F}$-predictable and bounded, the last term is an $\mathcal{F}$-martingale. Thus, we obtain

$$\begin{aligned}
v(t)W_{Z_t}(t) &= \mathrm{E}\bigg[v(t) \sum_{j\in S} \mathbf{1}_{\{Z_t=j\}} W_j(t) \,\bigg|\, \mathcal{F}_t \bigg] \\
&= \mathrm{E}\bigg[v(t) \int_{(t,n]} \frac{\kappa(t)}{\kappa(s)}\, B(\mathrm{d}s) \,\bigg|\, \mathcal{F}_t \bigg] = v(t)V(t)
\end{aligned}$$

almost surely for all $t \in [0,n]$. Noting $v > 0$ completes the proof. □

2.4.3 Examples

In this subsection, we consider the framework of stochastic retirement from Subsection 2.2.3 and the non-monotone information given by $\mathcal{G}^1$ and $\mathcal{G}^2$. The time of retirement and death are given by the hitting times η and δ, respectively. Recall that $\mathcal{G}^1$ corresponds to the case where upon retirement or death the insurer discards the previous health records of the insured, while $\mathcal{G}^2$ even keeps no record on the time of retirement.

In Subsection 2.4.3, we present some auxiliary results characterizing the relevant IF- and IB-compensators and state-wise quantities. Stochastic Thiele equations are then derived in Subsection 2.4.3 using the general theory developed in Subsection 2.4.2. Finally, in Subsection 2.4.3 we specialize the inter-temporal dependency structure, derive Feynman-Kac formulas, and relate the results to actuarial practice.

Preliminaries

Denote for $j,k \in S$, $k \neq j$, by $\Gamma^{\mathcal{G}^i_-}_{jk}$ and $\Gamma^{\mathcal{G}^i}_{jk}$ the IF- and IB-compensator of N_{jk}, respectively, w.r.t. $\mathcal{G}^i$, $i = 1,2$. Recall that Λ denotes the classic $\mathcal{F}$-compensators of N. The following result gives an explicit characterization of the relevant IF- and IB-compensators of N w.r.t. $\mathcal{G}^1$ and $\mathcal{G}^2$.

Proposition 2.4.7. *For all $t > 0$ we almost surely have*

$$\begin{aligned}
\Gamma^{\mathcal{G}^1-}_{jk}(t) &= \Gamma^{\mathcal{G}^2_-}_{jk}(t) = \Lambda_{jk}(t), \quad j \in \{1,\dots,J\},\, k \in S \setminus \{j\},\\
\Gamma^{\mathcal{G}^1}_{jk}(t) &= \Gamma^{\mathcal{G}^2}_{jk}(t) = N_{jk}(t), \quad j \in S,\, k \in \{1,\dots,J\} \setminus \{j\} \text{ or } j = J+1,\, k = J+2,\\
\Gamma^{\mathcal{G}^1-}_{jk}(t) &= \int_{(0,t]} \frac{\mathbf{1}_{\{\eta < s \leq \delta\}}}{P(\delta \geq s \,|\, \eta)} P(\delta \in \mathrm{d}s \,|\, \eta), \quad j = J+1,\, k = J+2,\\
\Gamma^{\mathcal{G}^2-}_{jk}(t) &= \int_{(0,t]} \frac{\mathbf{1}_{\{\eta < s \leq \delta\}}}{P(\eta < s \leq \delta)} P(\delta \in \mathrm{d}s, Z_{\delta-} = J+1), \quad j = J+1,\, k = J+2,\\
\Gamma^{\mathcal{G}^1}_{jk}(t) &= \int_{(0,t]} P(Z_{\eta-} = j \,|\, \eta = s) \sum_{\ell=1}^{J} N_{\ell k}(\mathrm{d}s), \quad j \in \{1,\dots,J\},\, k = J+1,\\
\Gamma^{\mathcal{G}^2}_{jk}(\mathrm{d}t) &= \mathbf{1}_{\{\eta \leq t < \delta\}} G^{\mathcal{G}^2}_{jk}(\mathrm{d}t), \quad j \in \{1,\dots,J\}, k = J+1,\\
G^{\mathcal{G}^2}_{jk}(t) &:= \int_{(0,t]} \frac{\mathbf{1}_{\{P(\eta \leq s < \delta) > 0\}}}{P(\eta \leq s < \delta)} P(\eta \in \mathrm{d}s, Z_{\eta-} = j), \quad j \in \{1,\dots,J\}, k = J+1,\\
\Gamma^{\mathcal{G}^1}_{jk}(t) &= \Gamma^{\mathcal{G}^2}_{jk}(t)\\
&= \int_{(0,t]} \frac{P(Z_{\delta-} = j \,|\, \delta = s)}{P(Z_{\delta-} \neq J+1 \,|\, \delta = s)} \sum_{\ell=1}^{J} N_{\ell k}(\mathrm{d}s), \quad j \in \{1,\dots,J\},\, k = J+2.
\end{aligned}$$

All remaining IF- and IB-compensators of N equal zero almost surely .

Sketch of proof. Calculate the IF-compensator $\gamma^{\mathcal{G}^i_-}_{xy}$ and the IB-compensator $\gamma^{\mathcal{G}^i_{xy}}$ of ν_{xy} from Proposition 2.4.1 and use the construction of $\mathcal{G}^1$ and $\mathcal{G}^2$ according to the proof of Lemma 2.2.4. □

In the following, $(Y^{\mathcal{G}^i}_j)_{j \in S}$ refers to the modification of the classic state-wise prospective reserves w.r.t. $\mathcal{G}^i$ presented in Example 2.3.15. The next result provides a characterization of the remaining key terms appearing in the stochastic Thiele equations w.r.t. $\mathcal{G}^1$ and $\mathcal{G}^2$.

Proposition 2.4.8. *For each $i \in \{1,2\}$ and $t > 0$ we have*

$$\begin{aligned}
b^i_{J+1}(t) &:= b^{\mathcal{G}^i_-}_{\{1\}}(t) = \mathrm{E}[b_{J+1}(t) \,|\, \mathcal{G}^i_{t-}],\\
\beta^i_{(J+1)(J+2)}(t) &:= \beta^{\mathcal{G}^i_-}_{\{1\}\{1,2\}}(t, (t, J+2)) = \mathrm{E}[b_{(J+1)(J+2)}(t) \,|\, \mathcal{G}^i_{t-}, \delta = t]
\end{aligned}$$

almost surely on $\{Z_{t-} = J+1\}$,

$$b^i_{J+2}(t) := \mathbf{1}_{\{\eta < t\}} b^{\mathcal{G}^i_-}_{\{1,2\}}(t) + \mathbf{1}_{\{\eta \geq t\}} b^{\mathcal{G}^i_-}_{\{2\}}(t) = \mathrm{E}[b_{J+2}(t) \,|\, \mathcal{G}^i_{t-}]$$

almost surely on $\{Z_{t-} = J+2\}$,

$$\begin{aligned}
R^i_{(J+1)(J+2)}(t) :=& \beta^i_{(J+1)(J+2)}(t) + Y^{\mathcal{G}^i}_{J+2}(t) - Y^{\mathcal{G}^i}_{\{1\}\{1\}}(t)\\
&= \mathrm{E}[\beta(t) + Y(t) \,|\, \mathcal{G}^i_{t-}, \delta = t] - \mathrm{E}[Y(t) \,|\, \mathcal{G}^i_{t-}, Z_t = J+1]
\end{aligned}$$

almost surely on $\{Z_{t-} = J+1\}$*, and for each* $t \geq 0$ *and* $j \in \{1, \dots, J\}$ *we have*

$$\begin{aligned} L^2_{j(J+1)}(t) &:= \mathrm{E}[Y(t) \,|\, \eta = t, Z_{\eta-} = j] - \mathcal{Y}^{\mathcal{G}^2}_{\{1\}\{1\}}(t) \\ &= \mathrm{E}[Y(t) \,|\, \mathcal{G}^2_t, \eta = t, Z_{\eta-} = j] - \mathrm{E}[Y(t) \,|\, \mathcal{G}^2_t, Z_{t-} = J+1] \end{aligned}$$

almost surely on $\{Z_t = J+1\}$.

Sketch of proof. Combine suitably the contents of Example 2.3.15, Proposition 2.3.7, the constructions of $\mathcal{G}^1$ and $\mathcal{G}^2$ according to the proof of Lemma 2.2.4, and (2.3.3). □

Stochastic Thiele equations

Based on the characterization of relevant IF- and IB-compensators and state-wise quantities from Subsection 2.4.3, the following two theorems yield stochastic Thiele equations for the classic state-wise prospective reserves w.r.t. non-monotone information $\mathcal{G}^1$ and $\mathcal{G}^2$.

Theorem 2.4.9. *The classic state-wise prospective reserves* $(Y_j^{\mathcal{G}^1})_{j \in S}$ *almost surely satisfy* $Y_j^{\mathcal{G}^1} = V_j$ *for* $j \in \{1, \dots, J\}$ *and*

$$\begin{aligned} 0 = \mathbf{1}_{\{Z_{t-}=J+1\}} \bigg(& Y^{\mathcal{G}^1}_{J+1}(\mathrm{d}t) - Y^{\mathcal{G}^1}_{J+1}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b^1_{J+1}(t)\, \mu(\mathrm{d}t) \\ & + R^1_{J+1(J+2)}(t)\, \Gamma^{\mathcal{G}^1_-}_{(J+1)(J+2)}(\mathrm{d}t) \bigg), \\ 0 = \mathbf{1}_{\{Z_{t-}=J+2\}} \bigg(& Y^{\mathcal{G}^1}_{J+2}(\mathrm{d}t) - Y^{\mathcal{G}^1}_{J+2}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b^1_{J+2}(t)\, \mu(\mathrm{d}t) \bigg). \end{aligned}$$

Theorem 2.4.10. *The classic state-wise prospective reserves* $(Y_j^{\mathcal{G}^2})_{j \in S}$ *almost surely satisfy* $Y_j^{\mathcal{G}^2} = V_j$ *for* $j \in \{1, \dots, J\}$ *and*

$$\begin{aligned} 0 = \mathbf{1}_{\{Z_{t-}=J+1\}} \bigg(& Y^{\mathcal{G}^2}_{J+1}(\mathrm{d}t) - Y^{\mathcal{G}^2}_{J+1}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b^2_{J+1}(t)\, \mu(\mathrm{d}t) \\ & + R^2_{J+1(J+2)}(t)\, \Gamma^{\mathcal{G}^2_-}_{(J+1)(J+2)}(\mathrm{d}t) \\ & - \sum_{k=1}^{J} L^2_{k(J+1)}(t)\, G^{\mathcal{G}^2}_{k(J+1)}(\mathrm{d}t) \bigg), \\ 0 = \mathbf{1}_{\{Z_{t-}=J+2\}} \bigg(& Y^{\mathcal{G}^2}_{J+2}(\mathrm{d}t) - Y^{\mathcal{G}^2}_{J+2}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b^2_{J+2}(t)\, \mu(\mathrm{d}t) \bigg). \end{aligned}$$

Sketch of proof of Theorem 2.4.9 and Theorem 2.4.10. Since $\{Z_t = J+1\} = \{\eta \leq t < \delta\} = \{\mathcal{Z}_t = \{1\}\}$, $\{Z_t = J+2, \eta \leq t\} = \{\mathcal{Z}_t = \{1,2\}\}$, and $\{Z_t = J+2, \eta > t\} = \{\mathcal{Z}_t = \{2\}\}$ for all $t \geq 0$, following along the lines of the proof of Corollary 2.4.5

and pointing to Example 2.3.15 yields

$$\begin{aligned}
\mathbf{1}_{\{Z_{t-}=J+1\}} Y^{\mathcal{G}^i}_{J+1}(\mathrm{d}t) &= I_{\{1\}}(t-)\mathcal{Y}^{\mathcal{G}^i}_{\{1\}}(\mathrm{d}t), \\
\mathbf{1}_{\{Z_{t-}=J+2\}} Y^{\mathcal{G}^i}_{J+2}(\mathrm{d}t) &= I_{\{1,2\}}(t-)\mathcal{Y}^{\mathcal{G}^i}_{\{1,2\}}(\mathrm{d}t) + I_{\{2\}}(t-)\mathcal{Y}^{\mathcal{G}^i}_{\{2\}}(\mathrm{d}t)
\end{aligned}$$

almost surely. Now apply Theorem 2.4.2, calculate the terms explicitly, collect them, and apply Proposition 2.4.7 and Proposition 2.4.8. □

Remark 2.4.11. Note that the term

$$\mathbf{1}_{\{Z_{t-}=J+1\}} \sum_{k=1}^{J} L^2_{k(J+1)}(t)\, G^{\mathcal{G}^2}_{k(J+1)}(\mathrm{d}t)$$

can be replaced by

$$\mathbf{1}_{\{Z_{t-}=J+1\}} L^2_{\bullet(J+1)}(t)\, G^{\mathcal{G}^2}_{\bullet(J+1)}(\mathrm{d}t),$$

where

$$\begin{aligned}
L^2_{\bullet(J+1)}(t) &:= \mathrm{E}[Y(t) \,|\, \eta = t] - Y^{\mathcal{G}^2}_{\{1\}\{1\}}(t) \\
&= \mathrm{E}[Y(t) \,|\, \eta = t] - \mathrm{E}[Y(t) \,|\, \eta < t < \delta]. \\
G^{\mathcal{G}^2}_{\bullet(J+1)}(\mathrm{d}t) &:= \frac{\mathbf{1}_{\{P(\eta \le t < \delta) > 0\}}}{P(\eta \le t < \delta)} P(\eta \in \mathrm{d}t)
\end{aligned}$$

almost surely. To see this, apply Proposition 2.4.7 and Proposition 2.4.8. ▽

The stochastic differential equations that follow from Theorem 2.4.9 and Theorem 2.4.10 are fundamentally different from the stochastic differential equations appearing in the presence of monotone information. Since $Y^{\mathcal{G}^i}_j$ almost surely equals V_j for $j \in \{1, \dots, J\}$, Corollary 2.4.5 yields the stochastic differential equations

$$\begin{aligned}
0 \overset{\text{a.s.}}{=} \mathbf{1}_{\{Z_{t-}=j\}} \Bigg(& Y^{\mathcal{G}^i}_j(\mathrm{d}t) - Y^{\mathcal{G}^i}_j(t-)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} + b_j(t)\,\mu(\mathrm{d}t) \\
& + \sum_{k:k\neq j} \big(b_{jk}(t) + V_k(t) - Y^{\mathcal{G}^i}_j\big)\,\Lambda_{jk}(\mathrm{d}t)\Bigg)
\end{aligned}$$

for $j \in \{1, \dots, J\}$. The sum at risks for $k \in \{J+1, J+2\}$ take an unusual form as they involve V_{J+1} and V_{J+2} rather than $Y^{\mathcal{G}^i}_{J+1}$ and $Y^{\mathcal{G}^i}_{J+2}$. Since information discarding occurs upon or after retirement and death, this just reflects full utilization of all available information (before retirement and death).

Another fundamental difference is evident in Theorem 2.4.10. Recall that $\mathcal{G}^2$ does not have the time since retirement as admissible information. Referring to Remark 2.4.11, the stochastic differential equation for $Y^{\mathcal{G}^2}_{J+1}$ includes the term

$$\Big(\mathrm{E}[Y(t) \,|\, \eta = t] - \mathrm{E}[Y(t) \,|\, \eta < t < \delta]\Big) \frac{\mathbf{1}_{\{P(\eta \le t < \delta) > 0\}}}{P(\eta \le t < \delta)} P(\eta \in \mathrm{d}t).$$

It adjusts the dynamics to take into account the possibility that retirement might just have occurred rather than having occurred some time ago (conditionally on the insured presently being retired). In the former case, at time t one would reserve $\mathrm{E}[Y(t) \mid \eta = t]$, while in the latter case one would reserve $E[Y(t) \mid \eta < t < \delta]$. This constitutes a description of the first part of the product. The second part is exactly the infinitesimal probability of retirement having just occurred, conditionally on the insured presently being retired.

Feynman-Kac formulas

We now specialize and simplify the setting to provide a more straightforward and less technical discussion of the general results and their relation to actuarial practice.

Suppose that $\tilde{Z}$ is semi-Markovian such that the $\mathcal{F}$-compensators Λ of N admit densities w.r.t. the Lebesgue measure and such that (η, δ) is a continuous random variable. Denote by $f_{(\eta,\delta)}$ the joint density function of (η, δ), by $f_{\eta|\delta}$ the conditional density function of η given δ, and by f_η and f_δ the marginal density functions of η and δ. Further, suppose that b_j and b_{jk} are deterministic for all $j, k \in S$, $j \neq k$, and let there be a maximal contract time $n < \infty$, i.e. each b_j and b_{jk} is constantly zero on the interval (n, ∞).

Because of Proposition 2.2.3, the compensators Λ have representations of the form (for $j \in \{1, \dots, J\}$, $k \in \{1, \dots, J+2\} \setminus \{j\}$)

$$\begin{aligned} \Lambda_{jk}(\mathrm{d}t) &= \mathbf{1}_{\{Z_{t-}=j\}} \alpha_{jk}(t, t - U_{t-}) \, \mathrm{d}t, \\ \Lambda_{(J+1)(J+2)}(\mathrm{d}t) &= \mathbf{1}_{\{Z_{t-}=J+1\}} \alpha_{(J+1)(J+2)}(t, t - U^h_{t-}, t - U^r_{t-}, H_{t-}) \, \mathrm{d}t \end{aligned}$$

for deterministic functions α_{jk} and $\alpha_{(J+1)(J+2)}$, so-called transition rates.

The next results provide Feynman-Kac formulas that can serve as the starting point for the development of numerical schemes for the classic state-wise prospective reserves $(V_j)_{j \in S}$.

Proposition 2.4.12. *Suppose the assumptions from the beginning of this subsection hold. If the function $W_{J+2}(\cdot)$ is a bounded càdlàg solution of*

$$W_{J+2}(\mathrm{d}t) = W_{J+2}(t-) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_{J+2}(t) \, \mu(\mathrm{d}t), \qquad t > 0, \tag{2.4.5}$$

with terminal condition $W_{J+2}(n) = 0$, and the function $W_{J+1}(\cdot, \cdot, \cdot, \cdot)$ is a bounded and càdlàg solution of (for $t > r > s \geq 0$, $k \in \{1, \dots, J\}$)

$$\begin{aligned} &W_{J+1}(\mathrm{d}t, s, r, k) \\ &= W_{J+1}(t-, s, r, k) \frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_{J+1}(t) \, \mu(\mathrm{d}t) \\ &\quad - \big(b_{(J+1)(J+2)}(t) + W_{J+2}(t) - W_{J+1}(t, s, r, k) \big) \, \alpha_{(J+1)(J+2)}(t, s, r, k) \, \mathrm{d}t, \end{aligned} \tag{2.4.6}$$

with terminal conditions $W_{J+1}(n,s,r,k) = 0$ *for* $0 \leq s \leq r \leq n$ *and* $k \in \{1,\dots,J\}$, *and the functions* $W_j(\cdot,\cdot)$, $j \in \{1,\dots,J\}$, *are bounded and càdlàg solutions of (for* $t > s \geq 0$, $j \in \{1,\dots,J\}$*)*

$$\begin{aligned} W_j(\mathrm{d}t, s) = W_j(t-, s)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} &- b_j(t)\,\mu(\mathrm{d}t) + \\ &- \sum_{k \leq J: k \neq j} \big(b_{jk}(t) + W_k(t,t) - W_j(t,s)\big)\,\alpha_{jk}(t,s)\,\mathrm{d}t \\ &- \big(b_{j(J+1)}(t) + W_{J+1}(t,s,t,j) - W_j(t,s)\big)\,\alpha_{j(J+1)}(t,s)\,\mathrm{d}t \\ &- \big(b_{j(J+2)}(t) + W_{J+2}(t) - W_j(t,s)\big)\,\alpha_{j(J+2)}(t,s)\,\mathrm{d}t, \end{aligned} \tag{2.4.7}$$

with terminal conditions $W_j(n,s) = 0$ *for* $0 \leq s \leq n$, *then for all* $t \geq 0$ *and* $j \in \{1,\dots,J\}$,

$$\mathbf{1}_{\{Z_t=j\}} W_j(t, t-U_t) = \mathbf{1}_{\{Z_t=j\}}\, V_j(t) = \mathbf{1}_{\{Z_t=j\}} V(t)$$

almost surely, and for all $t \geq 0$,

$$\begin{aligned} \mathbf{1}_{\{Z_t=J+1\}} W_{J+1}(t, t-U_t^h, t-U_t^r, H_t) &= \mathbf{1}_{\{Z_t=J+1\}} V_{J+1}(t) = \mathbf{1}_{\{Z_t=J+1\}} V(t), \\ \mathbf{1}_{\{Z_t=J+2\}} W_{J+2}(t) &= \mathbf{1}_{\{Z_t=J+2\}}\, V_{J+2}(t) = \mathbf{1}_{\{Z_t=J+2\}} V(t) \end{aligned}$$

almost surely.

Proof. Note that the right-continuity of the solutions of the differential/integral equations allows us to uniquely expand the domains of the solutions to $t \geq s \geq 0$, $t \geq r > s \geq 0$ and $t \geq 0$. That means that $W_j(t,t)$ and $W_{J+1}(t,s,t,k)$ are indeed given by the solutions.

Since the bounded and càdlàg solution W_{J+2} of (2.4.5) is deterministic, it is also $\mathcal{F}$-predictable and by multiplying (2.4.5) with $\mathbf{1}_{\{Z_{t-}=J+2\}}$ we obtain (2.4.4) for $j = J+2$. By multiplying equation (2.4.6) with $\mathbf{1}_{\{Z_{t-}=J+1\}}$ and replacing s, r and k by $t - U_{t-}^h$, $t - U_{t-}^r$, and H_{t-}, we obtain that $W_{J+1}(t, t-U_{t-}^h, t-U_{t-}^r, H_{t-})$ is an $\mathcal{F}$-predictable, bounded, and càdlàg solution of (2.4.4) for $j = J+1$.

Multiplying equation (2.4.7) with $\mathbf{1}_{\{Z_{t-}=j\}}\mathbf{1}_{\{\tau_i < t \leq \tau_{i+1}\}}$ and replacing s by $\tau_i \mathbf{1}_{\{Z_{\tau_i}=j\}} + t\,\mathbf{1}_{\{Z_{\tau_i}\neq j\}}$, we obtain that $W_j(t, \tau_i \mathbf{1}_{\{Z_{\tau_i}=j\}} + t\,\mathbf{1}_{\{Z_{\tau_i}\neq j\}})$, $j \in \{1,\dots,J\}$, is a solution of (2.4.4) on the interval $(\tau_i, \tau_{i+1}]$. This follows from the almost sure identities

$$\begin{aligned} \mathbf{1}_{\{Z_{t-}=j\}} W_k(t, \tau_i \mathbf{1}_{\{Z_{\tau_i}=k\}} + t\,\mathbf{1}_{\{Z_{\tau_i}\neq k\}}) &= \mathbf{1}_{\{Z_{t-}=j\}} W_k(t,t), \\ \mathbf{1}_{\{Z_{t-}=j\}} W_{J+1}(t, \tau_i \mathbf{1}_{\{Z_{\tau_i}=j\}} + t\,\mathbf{1}_{\{Z_{\tau_i}\neq j\}}, 0, j) & \\ &= \mathbf{1}_{\{Z_{t-}=j\}} W_{J+1}(t, t-U_{t-}^h, U_{t-}^r, H_{t-}) \end{aligned}$$

for all $t \in (\tau_i, \tau_{i+1}]$ and $j, k \in \{1, \dots, J\}$, $j \neq k$. Summing over $i \in \mathbb{N}_0$ yields that the bounded and càdlàg $\mathcal{F}$-predictable processes

$$W_j(t, t - U_{t-}\mathbf{1}_{\{Z_{t-}=j\}}) = \sum_{i=0}^{\infty} \mathbf{1}_{\{\tau_i < t \le \tau_{i+1}\}} W_j(t, \tau_i \mathbf{1}_{\{Z_{\tau_i}=j\}} + t\,\mathbf{1}_{\{Z_{\tau_i}\neq j\}})$$

are solutions of (2.4.4) for $j \in \{1, \dots, J\}$ due to the fact that

$$\mathbf{1}_{\{Z_{t-}=j\}} W_k(t, t - U_{t-}\mathbf{1}_{\{Z_{t-}=k\}}) = \mathbf{1}_{\{Z_{t-}=j\}} W_k(t, t)$$

almost surely for all $t > 0$ and $j, k \in \{1, \dots, J\}$, $j \neq k$.

All in all, we conclude that the processes $W_j(t, t - U_{t-}\mathbf{1}_{\{Z_{t-}=j\}})$, $j \in \{1, \dots, J\}$, $W_{J+1}(t, t - U^h_{t-}, t - U^r_{t-}, H_{t-})$, and $W_{J+2}(t)$ form an $\mathcal{F}$-predictable bounded and càdlàg solution of the equation system (2.4.4), which implies that, according to Proposition 2.4.6, they equal the classic state-wise prospective reserves $V_j(t)$ on $\{Z_t = j\}$ for $j \in \{1, \dots, J+2\}$. Since $Z_t = J+1$ implies $\eta \le t$, we may replace $\mathbf{1}_{\{Z_t=J+1\}} W_{J+1}(t, t - U^h_{t-}, t - U^r_{t-}, H_{t-})$ by $\mathbf{1}_{\{Z_t=J+1\}} W_{J+1}(t, t - U^h_t, t - U^r_t, H_t)$. Moreover, we have

$$\mathbf{1}_{\{Z_t=j\}} W_j(t, t - U_t) = \mathbf{1}_{\{Z_t=j\}} W_j(t, t - U_{t-}\mathbf{1}_{\{Z_{t-}=j\}}), \qquad j \in \{1, \dots, J\},$$

almost surely for all $t \ge 0$ under the conventions $U_{0-} := 0$ and $Z_{0-} := Z_0$. This implies the statement of the proposition. □

The numerical schemes that can be developed based on Proposition 2.4.12 are significantly more complex than in the classic (semi-)Markovian case, see e.g. Adékambi and Christiansen (2017). The sum at risks involve $W_{J+1}(t, s, t, j)$, which must be computed based on (2.4.6) for all $0 \le s < t$ using e.g. the method of lines.

Recall that $Y^{\mathcal{G}^i}_j = V_j$ almost surely for $j \in \{1, \dots, J\}$, cf. Theorem 2.4.9 and Theorem 2.4.10, and due to the assumptions given at the beginning of this subsection, we also have $Y^{\mathcal{G}^i}_{J+2} = V_{J+2}$ almost surely. The next results provide Feynman-Kac formulas for the residuary classic state-wise prospective reserve in the presence of non-monotone information $\mathcal{G}^1$ and $\mathcal{G}^2$. Proofs are given at the end of the subsection.

Proposition 2.4.13. *Suppose the assumptions from the beginning of this subsection hold. If $W^1_{J+1}(\cdot,\cdot)$ is a bounded and càdlàg solution of*

$$\begin{aligned} &W^1_{J+1}(\mathrm{d}t, r) \\ &= W^1_{J+1}(t-, r)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_{J+1}(t)\,\mu(\mathrm{d}t) \\ &\quad - \big(b_{(J+1)(J+2)}(t) + W_{J+2}(t) - W^1_{J+1}(t, r)\big)\,\alpha^1_{(J+1)(J+2)}(t, r)\,\mathrm{d}t \end{aligned} \tag{2.4.8}$$

for $0 < r < t$ with terminal conditions $W^1_{J+1}(n, r) = 0$ for $0 \le r \le n$, and where $W_{J+2}(\cdot)$ solves (2.4.5) while

$$\alpha^1_{(J+1)(J+2)}(t, r) := \frac{f_{\delta|\eta}(t|r)}{P(\delta \ge t \,|\, \eta = r)}, \qquad 0 \le r \le t, \tag{2.4.9}$$

then $\mathbf{1}_{\{Z_t=J+1\}} W^1_{J+1}(t,\eta) = \mathbf{1}_{\{Z_t=J+1\}} Y^{\mathcal{G}^1}_{J+1}(t) = \mathbf{1}_{\{Z_t=J+1\}} Y^{\mathcal{G}^1}(t)$ *almost surely for all* $t \geq 0$.

Proposition 2.4.14. *Suppose the assumptions from the beginning of this subsection hold. If* $W^2_{J+1}(\cdot,\cdot)$ *is a bounded and càdlàg solution of*

$$\begin{aligned} W^2_{J+1}(\mathrm{d}t) = {} & W^2_{J+1}(t-)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_{J+1}(t)\,\mu(\mathrm{d}t) \\ & - \big(b_{(J+1)(J+2)}(t) + W_{J+2}(t) - W^2_{J+1}(t)\big)\,\alpha^2_{(J+1)(J+2)}(t)\,\mathrm{d}t \\ & + \big(W^1_{J+1}(t,t) - W^2_{J+1}(t)\big)\,\xi_{J+1}(t)\,\mathrm{d}t \end{aligned} \tag{2.4.10}$$

for $0 < t$ *with terminal condition* $W^2_{J+1}(n) = 0$, *and where* $W_{J+2}(\cdot)$ *and* $W^1_{J+1}(\cdot,\cdot)$ *solve* (2.4.5) *and* (2.4.8) *while*

$$\alpha^2_{(J+1)(J+2)}(t) := \frac{\int_0^t f_{(\eta,\delta)}(s,t)\,\mathrm{d}s}{P(\eta < t \leq \delta)}, \tag{2.4.11}$$

$$\xi_{J+1}(t) := \frac{f_\eta(t)}{P(\eta \leq t < \delta)}, \tag{2.4.12}$$

then $\mathbf{1}_{\{Z_t=J+1\}} W^2_{J+1}(t) = \mathbf{1}_{\{Z_t=J+1\}} Y^{\mathcal{G}^2}_{J+1}(t) = \mathbf{1}_{\{Z_t=J+1\}} Y^{\mathcal{G}^2}(t)$ *almost surely for all* $t \geq 0$.

In order to reduce the computation time and simplify actuarial modeling and statistical estimation, practitioners, when computing the prospective reserve for non-retirees based on W_j, $j \in \{1,\dots,J\}$, often approximate $W_{J+1}(t,s,t,j)$ by a less complex quantity such as $W^1_{J+1}(t,t)$, which discards information concerning previous health records, or $W^2_{J+2}(t)$, which additionally discards information concerning the time of retirement. Replacing W_{J+1} by W^i_{J+1} produces approximation errors on the individual level (and redistribution of wealth on the portfolio level for non-retirees).

Proposition 2.4.13 and Proposition 2.4.14 can be used to develop computational schemes for W^1_{J+1} and W^2_{J+1}, respectively. Focusing on W^2_{J+1}, this involves the transition rate $\alpha^2_{(J+1)(J+2)}$, which by (2.4.11) is the hazard rate corresponding to a classic mortality table for retirees. It also involves the adjustment term

$$\big(W^1_{J+1}(t,t) - W^2_{J+1}(t)\big)\,\xi_{J+1}(t)\,\mathrm{d}t,$$

where according to (2.4.12), $\xi_{J+1}(t)\,\mathrm{d}t$ is the infinitesimal probability of retirement having just occurred (at time t), conditionally on the insured presently being retired.

If the mortality does not depend on the time since retirement, i.e. if

$$\alpha^1_{(J+1)(J+2)}(t,r) \equiv \alpha^2_{(J+1)(J+2)}(t),$$

we end up with the differential/integral equations

$$\begin{aligned}W^i_{J+1}(\mathrm{d}t) = {}& W^i_{J+1}(t-)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)} - b_{J+1}(t)\,\mu(\mathrm{d}t)\\ &- \big(b_{(J+1)(J+2)}(t) + W_{J+2}(t) - W^i_{J+1}(t)\big)\,\alpha^2_{(J+1)(J+2)}(t)\,\mathrm{d}t.\end{aligned} \tag{2.4.13}$$

Even though the mortality of retirees might depend on the time since retirement, practitioners often still utilize (2.4.13) directly. This produces additional approximation errors on the individual level (and redistribution of wealth on the portfolio level for retirees as well as non-retirees).

Proof of Proposition 2.4.13. Note that (2.4.8) implies that $W^1_{J+1}(\cdot,\eta)$ has paths of finite variation on compacts. By applying integration by parts, we obtain

$$\begin{aligned}&\mathbf{1}_{\{\eta<t\}}\,\mathrm{d}\Big(\mathbf{1}_{\{Z_t=J+1\}}v(t)W^1_{J+1}(t,\eta)\Big)\\ &= \mathbf{1}_{\{Z_{t-}=J+1\}}\Big(v(t)W^1_{J+1}(\mathrm{d}t,\eta) - v(t)W_{J+1}(t-,\eta)\frac{\kappa(\mathrm{d}t)}{\kappa(t-)}\Big)\\ &\quad - v(t)W^1_{J+1}(t,\eta)\,N_{(J+1)(J+2)}(\mathrm{d}t).\end{aligned}$$

almost surely. Inserting (2.4.8) into the latter term leads to

$$\begin{aligned}&\mathbf{1}_{\{\eta<t\}}\,\mathrm{d}\Big(\mathbf{1}_{\{Z_t=J+1\}}v(t)W^1_{J+1}(t,\eta)\Big)\\ &= -\mathbf{1}_{\{Z_{t-}=J+1\}}v(t)B(\mathrm{d}t) - v(t)W_{J+2}(t)N_{(J+1)(J+2)}(\mathrm{d}t)\\ &\quad + v(t)r^1_{(J+1)(J+2)}(t)M^1_{(J+1)(J+2)}(\mathrm{d}t)\end{aligned}$$

almost surely; here $r^1_{(J+1)(J+2)}(t) := b_{(J+1)(J+2)}(t) + W_{J+2}(t) - W^1_{J+1}(t,\eta)$ and $M^1_{(J+1)(J+2)}(\mathrm{d}t) := N_{(J+1)(J+2)}(\mathrm{d}t) - \mathbf{1}_{\{Z_{t-}=J+1\}}\alpha^1_{(J+1)(J+2)}(t,\eta)\,\mathrm{d}t$. Thus, since $\{\eta < t\} \subseteq \{\eta < s\}$ for $s \geq t \geq 0$, we find that almost surely for all $t \geq 0$,

$$\begin{aligned}&\mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)W^1_{J+1}(t,\eta)\\ &= \mathrm{E}[\mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)W^1_{J+1}(t,\eta)\,|\,\mathcal{G}^1_t]\\ &= \mathbf{1}_{\{\eta<t\}}\,\mathrm{E}\bigg[v(t)\int_{(t,n]}\mathbf{1}_{\{Z_{s-}=J+1\}}\frac{\kappa(t)}{\kappa(s)}\,B(\mathrm{d}s)\,\bigg|\,\mathcal{G}^1_t\bigg]\\ &\quad + \mathbf{1}_{\{\eta<t\}}\,\mathrm{E}\bigg[\int_{(t,n]}v(s)W_{J+2}(s)\,N_{(J+1)(J+2)}(\mathrm{d}s)\,\bigg|\,\mathcal{G}^1_t\bigg]\\ &\quad - \mathbf{1}_{\{\eta<t\}}\,\mathrm{E}\bigg[\int_{(t,n]}v(s)r^1_{(J+1)(J+2)}(s)\,M^1_{(J+1)(J+2)}(\mathrm{d}s)\,\bigg|\,\mathcal{G}^1_t\bigg]\\ &= \mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)Y^{\mathcal{G}^1}(t)\\ &\quad - \mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}\,\mathrm{E}\bigg[\int_{(t,n]}v(s)r^1_{(J+1)(J+2)}(s)\,M^1_{(J+1)(J+2)}(\mathrm{d}s)\,\bigg|\,\mathcal{G}^1_t\bigg]\end{aligned}$$

Recall that $\{Z_t = J+1\} = \{\mathcal{Z} \in \{1\}\}$. Pointing to Proposition 2.3.7, the constructions of $\mathcal{G}^1$ according to the proof of Lemma 2.2.4, and (2.3.3), straightforward calculations then yield that the last line equals zero. All in all, we conclude that

$$\begin{aligned}\mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)W^1_{J+1}(t,\eta) &= \mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)Y^{\mathcal{G}^1}(t)\\ &= \mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}v(t)Y^{\mathcal{G}^1}_{J+1}(t)\end{aligned}$$

almost surely for all $t \geq 0$. Since v and $Y^{\mathcal{G}^1}$ almost surely have càdlàg sample paths, cf. Proposition 2.3.8, we may replace $\mathbf{1}_{\{\eta<t\}}\mathbf{1}_{\{Z_t=J+1\}}$ by $\mathbf{1}_{\{\eta\leq t\}}\mathbf{1}_{\{Z_t=J+1\}} = \mathbf{1}_{\{Z_t=J+1\}}$. Using $v > 0$ completes the proof. □

Proof of Proposition 2.4.14. Since the distribution of η is assumed to admit a density w.r.t. the Lebesgue measure, we have $(0,\infty) \ni t \mapsto I_{\{1\}}(t) = I_{\{1\}}(t-)I_{\{1\}}(t)$ almost surely when

$$Y^{\mathcal{G}^2}_{J+1} = Y^{\mathcal{G}^2}_{\{1\}\{1\}}$$

almost surely, cf. Example 2.3.14 and (2.3.3). Note (2.4.10) implies that $W^2_{J+1}(\cdot)$ has paths of finite variation on compacts. By applying integration by parts, inserting (2.4.10), applying Theorem 2.4.10, and referring to Remark 2.4.11, straightforward calculations yield

$$\begin{aligned}&\mathrm{d}\Big(\mathbf{1}_{\{Z_t=J+1\}}v(t)W^2_{J+1}(t) - \mathbf{1}_{\{Z_t=J+1\}}v(t)\,Y^{\mathcal{G}^2}_{J+1}(t)\Big)^2\\ &= v(t)^2\mathbf{1}_{\{Z_{t-}=J+1\}}\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\big(\alpha^2_{(J+1)(J+2)}(t) - \xi_{J+1}(t)\big)\,\mathrm{d}t\\ &\quad - v(t)^2\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\big(N_{(J+1)(J+2)}(\mathrm{d}t) - \mathbf{1}_{\{Z_{t-}=J+1\}}\alpha^2_{(J+1)(J+2)}(t)\,\mathrm{d}t\big)\\ &\quad + v(t)^2\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\Big(-\mathbf{1}_{\{Z_t=J+1\}}\xi_{J+1}(t)\,\mathrm{d}t + \sum_{k=1}^{J}N_{k(J+1)}(\mathrm{d}t)\Big)\end{aligned}$$

almost surely. Following along the lines of the proof of Proposition 2.4.13, we find that

$$\begin{aligned}&v(t)^2P(Z_t = J+1)\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\\ &= \mathrm{E}\Big[\mathbf{1}_{\{Z_t=J+1\}}v(t)^2\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\Big]\\ &= -\,\mathrm{E}\bigg[\int_t^n v(s)^2\mathbf{1}_{\{Z_{s-}=J+1\}}\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\big(\alpha^2_{(J+1)(J+2)}(s) - \xi_{J+1}(s)\big)\,\mathrm{d}s\bigg]\\ &= -\int_t^n v(s)^2P(Z_s = J+1)\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2\big(\alpha^2_{(J+1)(J+2)}(s) - \xi_{J+1}(s)\big)\,\mathrm{d}s\end{aligned}$$

almost surely. This means that the function

$$f(t) := v(t)^2P(Z_t = J+1)\big(W^2_{J+1}(t) - Y^{\mathcal{G}^2}_{J+1}(t)\big)^2$$

almost surely satisfies the integral equation

$$f(t) = -\int_t^n f(s)\big(\alpha^2_{(J+1)(J+2)}(s) - \xi_{J+1}(s)\big)\,\mathrm{d}s$$

for all $t \in [0, n]$ under the convention $(n, n] = \emptyset$. Note that

$$|f(t)| \leq \int_t^n |f(s)| \left|\alpha^2_{(J+1)(J+2)}(s) - \xi_{J+1}(s)\right| \mathrm{d}s$$

almost surely for all $t \in [0, n]$. According to the the backward Grönwall inequality (see Cohen and Elliott, 2012, Lemma 4.7), $f(t) = 0$ almost surely for all $t \in [0, n]$. Since $v > 0$, for each $t \geq 0$ it then holds that $\mathbf{1}_{\{Z_t = J+1\}} W^2_{J+1}(t) = \mathbf{1}_{\{Z_t = J+1\}}\, Y^{\mathcal{G}^2}_{J+1}(t)$ almost surely. Since the implicated processes almost surely have càdlàg sample paths, cf. also with Example 2.3.15 and Proposition 2.3.8, there exists a joint P-null set. Thus $\mathbf{1}_{\{Z_t = J+1\}} W^2_{J+1}(t) = \mathbf{1}_{\{Z_t = J+1\}}\, Y^{\mathcal{G}^2}_{J+1}(t)$ almost surely for all $t \geq 0$ as desired. □

Acknowledgments and declarations of interest

Christian Furrer's research is partly funded by the Innovation Fund Denmark (IFD) under File No. 7038-00007B. The authors declare no competing interests.

2.A Proofs

Proof of Proposition 2.2.3. As a consequence of Assumption 2.2.2, the only non-trivial statement of the proposition relates to intertemporal dependency structure of Z after retirement, so it suffices to study the quantities

$$P(Z_s = J + 1 \,|\, \mathcal{F}_t\,)$$

on the event $\{Z_t = J + 1\}$ for $0 \leq t < s < \infty$. To this end, consider sets $A^n_t := \{Z_t = J + 1, \bar{N}(t) = n\}$, $n \in \mathbb{N}$, where $\bar{N} = (\bar{N}(t))_{t \geq 0}$ is the process counting the total number of jumps of Z given by

$$\bar{N}(t) = \sum_{\substack{j,k \in S \\ j \neq k}} N_{jk}(t), \qquad t \geq 0,$$

and denote with $\tau = (\tau_i)_{i \in \mathbb{N}}$ and $\tilde{\tau} = (\tilde{\tau}_i)_{i \in \mathbb{N}}$ the point processes corresponding to the jump times of Z and $\tilde{Z}$, respectively. Fix $0 \leq t < s < \infty$, and fix $n \in \mathbb{N}$. On A^n_t it then almost surely holds that

$$\begin{aligned} &\tau_n = \tilde{\tau}_n = \eta, && \tilde{Z}_{\tilde{\tau}_n} \in \{J + 1, \ldots, 2J\}, \\ &\tau_i = \tilde{\tau}_i && Z_{\tau_i} = \tilde{Z}_{\tilde{\tau}_i} \in \{1, \ldots, J\}, \qquad \forall i = 1, \ldots, n - 1. \end{aligned}$$

In particular,

$$\begin{aligned} &P(Z_s = J + 1 \,|\, \mathcal{F}_t\,)\mathbf{1}_{A^n_t} \\ &\overset{\text{a.s.}}{=} P\Big(\tilde{Z}_s \in \{J + 1, \ldots, 2J\} \,\Big|\, \tilde{\tau}_n, Z_{\tilde{\tau}_n}, \tilde{\tau}_{n-1}, \tilde{Z}_{\tilde{\tau}_{n-1}}, \ldots, \tilde{\tau}_1, \tilde{Z}_{\tilde{\tau}_1}\Big)\mathbf{1}_{A^n_t}. \end{aligned}$$

Suppose $(\tilde{Z},\tilde{U})$ is Markovian such that $\tilde{Z}$ is semi-Markovian. By the law of iterated expectations and the strong Markov property, cf. Theorem 7.5.1 in Jacobsen (2006), it follows that

$$\begin{aligned}
&P\Big(\tilde{Z}_s \in \{J+1,\ldots,2J\}\,\Big|\,\tilde{\tau}_n, Z_{\tilde{\tau}_n}, \tilde{\tau}_{n-1}, \tilde{Z}_{\tilde{\tau}_{n-1}},\ldots,\tilde{\tau}_1,\tilde{Z}_{\tilde{\tau}_1}\Big)\mathbf{1}_{A_t^n} \\
&= \mathrm{E}\Big[P\Big(\tilde{Z}_s \in \{J+1,\ldots,2J\}\,\Big|\,\tilde{\tau}_n, \tilde{Z}_{\tilde{\tau}_n}, \tilde{\tau}_n-\tilde{\tau}_{n-1}\Big)\,\Big|\,\tilde{\tau}_n, Z_{\tilde{\tau}_n},\ldots,\tilde{\tau}_1,\tilde{Z}_{\tilde{\tau}_1}\Big]\mathbf{1}_{A_t^j} \\
&\overset{\text{a.s.}}{=} \mathrm{E}\Big[P\Big(\tilde{Z}_s \in \{J+1,\ldots,2J\}\,\Big|\,\tilde{\tau}_n, \tilde{Z}_{\tilde{\tau}_n}, \tilde{\tau}_n-\tilde{\tau}_{n-1}\Big)\,\Big|\,\tilde{\tau}_n, Z_{\tilde{\tau}_n},\tilde{\tau}_{n-1},\tilde{Z}_{\tilde{\tau}_{n-1}}\Big]\mathbf{1}_{A_t^j} \\
&= P\Big(\tilde{Z}_s \in \{J+1,\ldots,2J\}\,\Big|\,\tilde{\tau}_n, Z_{\tilde{\tau}_n}, \tilde{\tau}_{n-1}, \tilde{Z}_{\tilde{\tau}_{n-1}}\Big)\mathbf{1}_{A_t^n}.
\end{aligned}$$

Thus on $A_t^n = \{Z_t = J+1, \bar{N}(t) = n\}$ it almost surely holds that

$$\begin{aligned}
P(Z_s = J+1 \,|\, \mathcal{F}_t) &= P\Big(Z_s = J+1\,\Big|\,\tilde{\tau}_n, Z_{\tilde{\tau}_n}, \tilde{\tau}_{n-1}, \tilde{Z}_{\tilde{\tau}_{n-1}}\Big) \\
&= P\Big(Z_s = J+1\,\Big|\,t-\tilde{\tau}_n, Z_{\tilde{\tau}_n}, t-\tilde{\tau}_{n-1}, \tilde{Z}_{\tilde{\tau}_{n-1}}\Big) \\
&= P\big(Z_s = J+1\,\big|\,U_t^r, Z_t, U_t^h, H_t\big),
\end{aligned}$$

which does not depend on n. We conclude that if $(\tilde{Z},\tilde{U})$ is Markovian, then on $\{Z_t = J+1\}$,

$$P(Z_s = J+1 \,|\, \mathcal{F}_t) \overset{\text{a.s.}}{=} P\big(Z_s = J+1\,\big|\,U_t^r, Z_t, U_t^h, H_t\big)$$

proving the first part of the proposition. The proof of the second and final part follows by similar arguments. ☐

Proof of Lemma 2.2.4. Let $N^- = (N^-(t))_{t\geq 0}$ be the process counting the number of jumps of Z except retirement and death given by

$$N^-(t) = \sum_{\substack{j,k\in S \\ k\notin\{j,J+1,J+2\}}} N_{jk}(t), \qquad t \geq 0,$$

and denote by $(\tau_i^-)_{i\in\mathbb{N}}$ the point process corresponding to the jumps of N^-. The σ-algebras (2.2.1) and (2.2.2) are equivalent to $\mathcal{G}_t^1$ and $\mathcal{G}_{t-}^1$, respectively, if we set

$$\begin{aligned}
T_1 &= \eta, & S_1 &= \infty, & \zeta_1 &= (T_1, Z_{T_1}), \\
T_2 &= \delta, & S_2 &= \infty, & \zeta_2 &= (T_2, Z_{T_2}), \\
T_{2+i} &= \tau_i^-, & S_{2+i} &= T_1 \wedge T_2, & \zeta_{2+i} &= (T_{2+i}, Z_{T_{2+i}}), \qquad i \in \mathbb{N}.
\end{aligned}$$

If we replace $\zeta_1 = (T_1, Z_{T_1})$ by the constant $\zeta_1 = (0, Z_{\tau_1}) = (0, J+1)$, then (2.2.1) and (2.2.2) are equivalent to $\mathcal{G}_t^2$ and $\mathcal{G}_{t-}^2$, respectively. ☐

Proof of Proposition 2.3.5. Since Y is integrable, for each $j \in S$ and $t \geq 0$ the mapping

$$C_{t,j} \ni A \mapsto \nu_{t,j}(A) := \int_A Y(t)\,\mathrm{d}m_{t,j}$$

is a finite signed measure on $\mathcal{C}_{t,j}$ which is absolutely continuous with respect to the sub-probability measure $m_{t,j}$ given by

$$C_{t,j} \ni A \mapsto m_{t,j}(A) = P(A \cap \{Z_t = j\}).$$

According to the Radon-Nikodym theorem there exist mappings $\omega \mapsto Y_j(t)(\omega)$ that are $\mathcal{C}_{t,j}$-measurable and satisfy

$$\nu_{t,j}(A) = \int_A Y_j(t)\,\mathrm{d}m_{t,j}, \qquad A \in \mathcal{C}_{t,j}. \tag{2.A.1}$$

In particular

$$\int_{A\cap\{Z_t=j\}} Y(t)\,\mathrm{d}P = \int_{A\cap\{Z_t=j\}} Y_j(t)\,\mathrm{d}P, \qquad j \in S, A \in \mathcal{C}_{t,j},$$

which by Lemma 2.3.1 yields

$$\int_A Y(t)\mathbf{1}_{\{Z_t=j\}}\,\mathrm{d}P = \int_A Y_j(t)\mathbf{1}_{\{Z_t=j\}}\,\mathrm{d}P, \qquad j \in S, A \in \mathcal{C}_t.$$

We conclude that $\mathbf{1}_{\{Z_t=j\}}Y_j(t) \overset{\text{a.s.}}{=} \mathbf{1}_{\{Z_t=j\}}Y(t)$ for each $j \in S$ and $t \geq 0$. This establishes existence of the state-wise counterparts. Furthermore, if there is another real-valued random variable $\tilde{Y}_j(t)$ that has the properties of $Y_j(t)$, we necessarily have

$$0 = \int_{A\cap\{Z_t=j\}} (Y_j(t) - \tilde{Y}_j(t))\,\mathrm{d}P = \int_{A\times\{j\}} (Y_j(t)(\omega) - \tilde{Y}_j(t)(\omega))\,\mathrm{d}m_t(\omega, j),$$

for $A \in \mathcal{C}_{t,j}$, which means that the mapping $(\omega, j) \mapsto Y_j(t)(\omega) - \tilde{Y}_j(t)(\omega)$ is m_t-almost everywhere zero. This establishes the desired uniqueness of the state-wise counterparts. □

Proof of Lemma 2.3.6. If $P(Z_t = j) = 0$, the result is trivial. Thus suppose $P(Z_t = j) > 0$. Since $\mathrm{E}_{t,j}[\mathrm{E}[X \,|\, \mathcal{C}_t] \,|\, \mathcal{C}_{t,j}]$ is the conditional expectation of $\mathrm{E}[X \,|\, \mathcal{C}_t]$ given $\mathcal{C}_{t,j}$ w.r.t. $P_{t,j}$, we find for $A \in \mathcal{C}_{t,j}$ that

$$\int_A \mathrm{E}_{t,j}[\mathrm{E}[X \,|\, \mathcal{C}_t] \,|\, \mathcal{C}_{t,j}]\;\mathrm{d}P_{t,j} = \int_A \mathrm{E}[X \,|\, \mathcal{C}_t]\;\mathrm{d}P_{t,j}.$$

Note that by definition of $\mathcal{C}_{t,j}$, we have $A \cap \{Z_t = j\} \in \mathcal{C}_t$. It follows that

$$\begin{aligned}
\int_A \mathrm{E}_{t,j}[\mathrm{E}[X \,|\, \mathcal{C}_t] \;|\, \mathcal{C}_{t,j}]\;\mathrm{d}P_{t,j} &= \frac{1}{P(Z_t = j)} \int_{A\cap\{Z_t=j\}} E[X \,|\, \mathcal{C}_t]\;\mathrm{d}P \\
&= \frac{1}{P(Z_t = j)} \int_A X\mathbf{1}_{\{Z_t=j\}}\,\mathrm{d}P \\
&= \int_A X\,\mathrm{d}P_{t,j},
\end{aligned}$$

where we have used that $\mathrm{E}[X \,|\, \mathcal{C}_t]$ is the conditional expectation of X given $\mathcal{C}_t$ w.r.t. P. In conclusion, $\mathrm{E}_{t,j}[\mathrm{E}[X \,|\, \mathcal{C}_{t,j}] \;|\, \mathcal{C}_{t,j}]$ is a version of the conditional expectation of X given $\mathcal{C}_{t,j}$ w.r.t. $P_{t,j}$ which completes the proof. □

Proof of Proposition 2.3.10. Suppose that $(T_i, S_i, \zeta_i)_{i\in\mathbb{N}} := (\tau_i, \infty, (Z_{\tau_i}, \tau_i))_{i\in\mathbb{N}}$, such that $\mathcal{F} = \mathcal{G}$. Fix $t > 0$ and $j \in S$. By (2.3.5) we almost surely find

$$\begin{aligned}
&\sum_{\substack{k\in S\\ j\neq k}} Y_{kj}^{\mathcal{F}-}(t)\\
&= \mathbf{1}_{\{Z_{t-}\neq j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \operatorname{E}[Y(t) \,|\, (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_{n+1}}, \tau_{n+1}) = (j,t)]\\
&= \mathbf{1}_{\{Z_{t-}\neq j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \frac{\operatorname{E}[\mathbf{1}_{\{(Z_{\tau_{n+1}}, \tau_{n+1}) = (j,t)\}} Y(t) \,|\, (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}{\operatorname{E}[\mathbf{1}_{\{(Z_{\tau_{n+1}}, \tau_{n+1}) = (j,t)\}} \,|\, (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}.
\end{aligned}$$

Since $\{Z_{t-} \neq j, Z_t = j, \tau_n < t \leq \tau_{n+1}\} = \{Z_{t-} \neq j, \tau_n < t \leq \tau_{n+1}, \tau_{n+1} = t, Z_{\tau_{n+1}} = j\}$ for any $n \in \mathbb{N}_0$, we further conclude on the basis of Example 2.3.2 and (2.3.2) that

$$\begin{aligned}
\sum_{\substack{k\in S\\ j\neq k}} Y_{kj}^{\mathcal{F}-}(t) &= \mathbf{1}_{\{Z_{t-}\neq j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \frac{\operatorname{E}[\mathbf{1}_{\{Z_t = j\}} Y(t) \,|\, (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}{\operatorname{E}[\mathbf{1}_{\{Z_t = j\}} \,|\, (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}\\
&= \mathbf{1}_{\{Z_{t-}\neq j\}} \sum_{n=0}^{\infty} \mathbf{1}_{\{\tau_n < t \leq \tau_{n+1}\}} \operatorname{E}[Y(t) \,|\, \mathcal{F}_{t-}, Z_t = j]\\
&= \mathbf{1}_{\{Z_{t-}\neq j\}} Y_j^{\mathcal{F}}(t)
\end{aligned}$$

almost surely. Similarly, $Y_{jj}^{\mathcal{F}-}(t) \overset{\text{a.s.}}{=} \mathbf{1}_{\{Z_{t-} = j\}} Y_j^{\mathcal{F}}(t)$. Writing

$$Y_j^{\mathcal{F}}(t) = Y_j^{\mathcal{F}}(t)\mathbf{1}_{\{Z_{t-}=j\}} + Y_j^{\mathcal{F}}(t)\mathbf{1}_{\{Z_{t-}\neq j\}}$$

and collecting terms completes the proof. □

Proof of Proposition 2.3.11. In this proof we generally suppose that $Z_{\tau_n} = j$.

The value of $Y_j^{\mathcal{F}}(t)$ at $t = r \wedge \tau_{n+1}$ is irrelevant for the càdlàg and finite variation path property. For $t \in (\tau_n, \tau_{n+1})$ we have $I_x(t-) = I_x(t-)I_x(t)$ and $I_x(t-)I_x(t)\mathcal{Y}_{xx}^{\mathcal{F}}(t) = I_x(t-)I_x(t)\mathcal{Y}_x^{\mathcal{F}}(t)$ because of (2.3.3) and (2.2.5). The latter fact and (2.3.4) yield

$$Y_j^{\mathcal{F}}(t) = Y_{jj}^{\mathcal{F}-}(t) = \mathbf{1}_{\{Z_{t-}=j\}} \sum_{x\in\mathcal{S}} I_x(t-)I_x(t)\mathcal{Y}_x^{\mathcal{F}}(t), \quad \tau_n < t < \tau_{n+1}, Z_t = j.$$

According to Proposition 2.3.8 the process $I_x(t-)I_x(t)\mathcal{Y}_x^{\mathcal{F}}$ has càdlàg paths of finite variation on $[0,r]$, so the same path proprties apply for $Y_j^{\mathcal{F}}$ on $[0,r] \cap (\tau_n, \tau_{n+1})$.

The value of $Y_j^{\mathcal{F}}(t)$ at $t = \tau_n$ is irrelevant for the finite variation path property, but it is relevant for the càdlàg property. By simplifying the second line of (2.3.5) to

$$Y_{jj}^{\mathcal{F}-}(t) = \frac{\operatorname{E}[\mathbf{1}_{\{\tau_{n+1} > t\}} Y(t) | (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}{\operatorname{E}[\mathbf{1}_{\{\tau_{n+1} > t\}} | (Z_{\tau_1}, \tau_1), \dots, (Z_{\tau_n}, \tau_n)]}, \quad \tau_n < t < \tau_{n+1}, Z_t = j,$$

and applying the Dominated Convergence Theorem, we obtain on $\{Z_{\tau_n} = j\}$ that

$$\begin{aligned}
\lim_{h \downarrow 0} Y_j^{\mathcal{F}}(\tau_n + h) &= \lim_{h \downarrow 0} Y_{jj}^{\mathcal{F}_-}(\tau_n + h) \\
&= \lim_{h \downarrow 0} \frac{\mathrm{E}[\mathbf{1}_{\{\tau_{n+1} > \tau_n + h\}} Y(\tau_n + h) | (Z_{\tau_1}, \tau_1), \ldots, (Z_{\tau_n}, \tau_n)]}{\mathrm{E}[\mathbf{1}_{\{\tau_{n+1} > \tau_n + h\}} | (Z_{\tau_1}, \tau_1), \ldots, (Z_{\tau_n}, \tau_n)]} \\
&= \mathrm{E}[Y(\tau_n) | (Z_{\tau_1}, \tau_1), \ldots, (Z_{\tau_n}, \tau_n)] \\
&= \sum_{\substack{k \in S \\ k \neq j}} Y_{kj}^{\mathcal{F}_-}(\tau_n) \\
&= Y_j^{\mathcal{F}}(\tau_n)
\end{aligned}$$

due to $\mathbf{1}_{\{\tau_{n+1} > \tau_n\}} = 1$, the first line of (2.3.5), and Proposition 2.3.10. □

Chapter 3

Computation of bonus in multi-state life insurance

This chapter contains the manuscript *Ahmad, Buchardt, and Furrer (2020).*

Abstract

We consider computation of market values of bonus payments in multi-state with-profit life insurance. The bonus scheme consists of additional benefits bought according to a dividend strategy that depends on the past realization of financial risk, the current individual insurance risk, the number of additional benefits currently held, and so-called portfolio-wide means describing the shape of the insurance business. We formulate numerical procedures that efficiently combine simulation of financial risk with more analytical methods for the outstanding insurance risk. Special attention is given to the case where the number of additional benefits bought only depends on the financial risk.

Keywords: Market consistent valuation; With-profit life insurance; Participating life insurance; Economic scenarios; Portfolio-wide means

3.1 Introduction

The potential of systematic surplus in multi-state with-profit life insurance (sometimes referred to as participating life insurance) leads to bonus payments that depend on the development of the financial market and the states of the insured. This dependence is typically non-linear and involves the whole paths of the processes governing the financial market and the states of the insured. Consequently, the computation of market values of bonus payments lies outside the scope of classic backward and forward methods. In this paper, we present computational schemes for a selection of these more involved market values using a combined approach in which we simulate the financial risk while retaining more analytical methods for the outstanding insurance risk.

In Denmark, the investment strategy and dividend strategy are to a great extent controlled by the insurer, and practitioners have traditionally determined the market value of bonus payments residually by imposing the equivalence principle on the market basis, cf. Møller and Steffensen (2007, Chapter 2). In reality, this valuation method is only applicable if – among other things – one includes payments to and from the equity, since such payments appear naturally in the context of e.g. cost of capital and other expenses. Thus a decomposition of the total market value that specifically displays the market value of bonus payments, as required by the Solvency II and IFRS 17 regulative frameworks, cf. EIOPA (2009, 2015) and IFRS (2017), cannot be derived residually unless the market value of payments to and from the equity is easy to determine. Since the latter generally is not the case, more sophisticated computational methods are required. The provision of these kinds of methods constitutes the main contribution of this paper.

The study of systematic surplus and bonus payments in multi-state with-profit life insurance goes back to Ramlau-Hansen (1991) and Norberg (1999, 2001), where one finds careful definitions of various concepts of surplus, discussions of general principles for its redistribution, and the introduction of forecasting techniques in a so-called Markov chain interest model, see also Norberg (1995). In Steffensen (2006), partial differential equations for market values of so-called predetermined payments and bonus payments are derived in a Black-Scholes model.

The projection of bonus payments in multi-state life insurance and the computation of associated market values has recently received renewed attention, see Jensen and Schomacker (2015), Jensen (2016), Bruhn and Lollike (2020), and Falden and Nyegaard (2020). In Jensen (2016), the focus is on projection of bonus payments conditionally on the insured sojourning in a specific state; this approach targets e.g. product design and bonus prognosis from the perspective of the insured rather than market valuation. Conversely, the paper Jensen and Schomacker (2015) also deals with projection of bonus payments but on a portfolio level, which ensures computational feasibility but does not shed light on the full complexity of multi-state with-profit life insurance. Although with-profit life insurance focuses on the collective and although decisions by the insurer (so-called future management actions), including possible determination of dividend yields, often depend mainly on the performance of the collective, one ought to take into account that bonus payments are individual in nature. This is the starting point in Bruhn and Lollike (2020), where the focus is on deriving differential equations for relevant retrospective reserves given a dividend strategy (used to buy additional benefits) that depends in an affine manner on the reserves themselves. The process governing the state of the insured is assumed Markovian. In Falden and Nyegaard (2020), the results of Bruhn and Lollike (2020) are extended to allow for policyholder behavior, namely the options of surrender and free policy conversion. In Bruhn and Lollike (2020) and Falden and Nyegaard (2020), the dependence of the dividend strategy on the performance of the collective, encapsulated in what we shall term the *shape* of the insurance business, and the practical and computational challenges arising from this are not highlighted.

In this paper, we derive methods for the computation of market values of bonus payments in a Markovian multi-state model for a financial market consisting of one risky asset in addition to a bank account governed by a potentially stochastic interest rate. The insurance risk and financial risk are assumed independent. We include the policyholder options surrender and free policy conversion following Henriksen et al. (2014), Buchardt and Møller (2015), and Buchardt, Møller, and Schmidt (2015) and focus on the bonus scheme known as *additional benefits*, where dividends are used to buy extra benefits; this bonus scheme is common in practice and is e.g. the focal point of Møller and Steffensen (2007, Chapter 2).

In practice, the dividend strategy depends on product design, regulatory frameworks, and decisions made by the insurer. In this paper, we assume that the dividend strategy is explicitly computable based on the following information: the past realization of financial risk, the current individual insurance risk (state of insured and time since free policy conversion), the current shape of the insurance business, and the number of additional benefits currently held. Furthermore, the dividend strategy must be affine in the number of additional benefits. The shape of the insurance business consists of so-called portfolio-wide means, cf. Møller and Steffensen (2007, Chapter 6), which reflect on a portfolio level the current financial state of the insurance business. Consequently, the shape of the insurance business depends on the dividend strategy, which again depends on the shape of the insurance business.

Using classic techniques, we derive a system of differential and integral equations for the computation of the expected accumulated bonus cash flows conditionally on the realization of financial risk. This allows us to formulate a procedure for the computation of the market value of bonus payments which efficiently combines simulation of financial risk with classic methods for the remaining insurance risk. We identify the special case where the number of additional benefits depend only on financial risk – the *state independent* case – and show how this significantly simplifies the numerical procedure. It is our impression that the state independent model is aligned to current actuarial practice, where it might e.g. serve as an approximation for valuation on a portfolio level.

We should like to stress that while our results are subject to important technical regularity conditions, it is the general methodology and conceptual ideas that constitute the main contributions of this paper. Furthermore, our concepts, methods, and results are targeted academics and actuarial practitioners alike, and, consequently, we aim at keeping the presentation at a reasonable technical level.

The paper is structured as follows. In Section 3.2, we present the setup. The general results and general numerical procedure are given in Section 3.3, while the state independent case is the subject of Section 3.4. Finally, Section 3.5 concludes with a comparison with recent advances in the literature and a discussion of possible extensions.

3.2 Setup

In the following, we describe the mathematical framework. Subsections 3.2.1–3.2.3 introduce the processes governing the financial market, the state of the insured, and the insurance payments, and we discuss the valuation of so-called predetermined payments. The dividend and bonus scheme is described in Subsection 3.2.4, which leads to a specification of the total payment stream as a sum of predetermined payments and bonus payments. Contrary to the predetermined payments, the bonus payments depend on the development of the financial market, which adds an extra layer of complexity to the valuation problem. The focal point of this paper is to establish explicit methods for the computation of the market value of the bonus payments; a precise description of this problem is given in Subsection 3.2.5. In the remainder of the paper, the problem is studied for a specific class of dividend processes specified in Subsection 3.2.6.

A background probability space $(\Omega, \mathbb{F}, \mathrm{P})$ is taken as given. Unless explicitly stated or evident from the specific context, all statements are in an almost sure sense w.r.t. P. The probability measure P relates to market valuation and therefore corresponds to some risk neutral probability measure. Due to the presence of insurance risk, the market is not complete, which implies that the risk neutral probability measure is not unique. Since we shall assume financial risk and insurance risk to be independent, one can think of the probability measure P as the product measure of some risk neutral probability measure for financial risk and some probability measure for insurance risk.

3.2.1 Preliminaries

The state of the insured is governed by a non-explosive jump process $Z = \{Z(t)\}_{t\geq 0}$ on a finite state space $\mathcal{J}$ with deterministic initial state $Z(0) \equiv z_0 \in \mathcal{J}$. Denote by N the corresponding multivariate counting process with components $N_{jk} = \{N_{jk}(t)\}_{t\geq 0}$ for $j, k \in \mathcal{J}, k \neq j$ given by

$$N_{jk}(t) = \#\{s \in (0, t] : Z(s-) = j,\ Z(s) = k\}.$$

Let $S_1 = \{S_1(t)\}_{t \geq 0}$ be the price process for some risky asset (diffusion process, in particular continuous) and let $r = \{r(t)\}_{t \geq 0}$ be a suitably regular short rate process with corresponding bank account $S_0(t) = S_0(0) \exp\left(\int_0^t r(v)\,\mathrm{d}v\right)$, $S_0(0) \equiv s_0 > 0$, and suitably regular forward interest rates $f(t, \cdot)$, $t \geq 0$, satisfying

$$\mathrm{E}\left[e^{-\int_t^T r(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}^S(t)\right] = e^{-\int_t^T f(t,s)\,\mathrm{d}s}$$

for all $0 \leq t < T$ as well as $f(t,t) = r(t)$ for all $t \geq 0$; here $\mathcal{F}^S$ is the natural filtration generated by $S := (S_0, S_1)$, which exactly represents available market information. The available insurance information is represented by the filtration $\mathcal{F}^Z$ naturally generated by Z, and the total information available is represented by the filtration $\mathcal{F} = \mathcal{F}^S \vee \mathcal{F}^Z$ naturally generated by (S, Z).

To allow for free policy behavior and surrender, we suppose the state space $\mathcal{J}$ can be decomposed as

$$\mathcal{J} = \mathcal{J}^{\mathrm{p}} \cup \mathcal{J}^{\mathrm{f}},$$

with $\mathcal{J}^{\mathrm{p}} := \{0, \ldots, J\}$ and $\mathcal{J}^{\mathrm{f}} := \{J+1, \ldots, 2J+1\}$ for some $J \in \mathbb{N}$. Here $\mathcal{J}^{\mathrm{p}}$ contains the premium paying states, while $\mathcal{J}^{\mathrm{f}}$ contains the free policy states, and transition to $\{J\}$ and $\{2J+1\}$ corresponds to surrender as premium paying and free policy, respectively, cf. Buchardt and Møller (2015) and Buchardt, Møller, and Schmidt (2015). We suppose that $\mathcal{J}^{\mathrm{f}}$ is absorbing and can only be reached via a transition from $\{0\}$ to $\{J+1\}$, $\{J\}$ and $\{2J+1\}$ are absorbing, and that $\{J\}$ and $\{2J+1\}$ can only be reached from $\{0\}$ and $\{J+1\}$, respectively. The setup is depicted in Figure 3.1.

3.2.2 Life insurance contract with policyholder options

The life insurance contract is described by a payment stream $B = \{B(t)\}_{t \geq 0}$ giving accumulated benefits less premiums. It consists of *predetermined payments* $B^{\circ} = \{B^{\circ}(t)\}_{0 \leq t \leq n}$, stipulated from the beginning of the contract, and additional bonus payments determined when market and insurance information are realized during the course of the contract; details regarding the latter are given in later subsections.

We specify the predetermined payments as in Buchardt and Møller (2015) and Buchardt, Møller, and Schmidt (2015). For simplicity, we suppose that the predetermined payments regarding the classic states $\mathcal{J}^{\mathrm{p}}$ consist of suitably regular deterministic sojourn payment rates b_j and transition payments b_{jk}; in particular, surrender results in a deterministic payment. In the free policy states, no premiums are paid and the benefit payments are reduced by a factor $\rho \in [0, 1]$ depending on

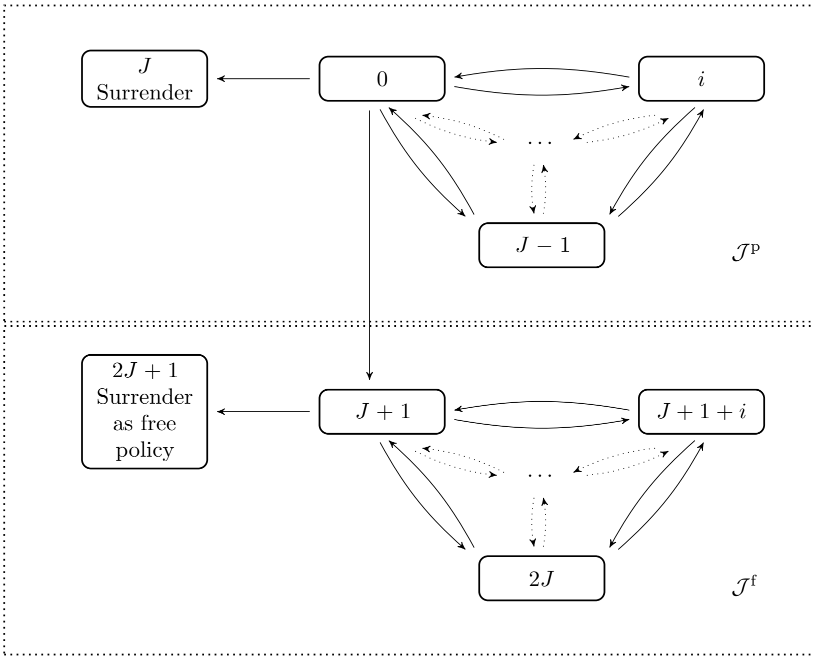

Figure 3.1: *General finite state space extended with a surrender state $\{J\}$ and free policy states $\mathcal{J}^{\mathrm{f}}$. The states $\mathcal{J}^{\mathrm{p}} \setminus \{J\}$ contain the biometric states of the insured, e.g. active, disabled, and dead. The states $\mathcal{J}^{\mathrm{f}}$ are a copy of $\mathcal{J}^{p}$, and a transition from $\{0\}$ to $\{J+1\}$ corresponds to a free policy conversion. A transition to $\{J\}$ or $\{2J+1\}$ corresponds to a surrender of the policy.*

the time of free policy conversion. In rigorous terms, we have

$$
\begin{aligned}
\mathrm{d}B^{\circ}(t) &= \mathrm{d}B^{\circ,\mathrm{p}}(t) + \rho(\tau)\,\mathrm{d}B^{\circ,\mathrm{f}}(t), && B^{\circ}(0) = 0,\\
\mathrm{d}B^{\circ,\mathrm{p}}(t) &= \sum_{j \in \mathcal{J}^{\mathrm{p}}} \mathbb{1}_{(Z(t-)=j)} \Bigg(b_j(t)\,\mathrm{d}t + \sum_{\substack{k \in \mathcal{J}^{\mathrm{p}} \\ k \neq j}} b_{jk}(t)\,\mathrm{d}N_{jk}(t) \Bigg), && B^{\circ,\mathrm{p}}(0) = 0,\\
\mathrm{d}B^{\circ,\mathrm{f}}(t) &= \sum_{j \in \mathcal{J}^{\mathrm{f}}} \mathbb{1}_{(Z(t-)=j)} \Bigg(b_{j'}^{+}(t)\,\mathrm{d}t + \sum_{\substack{k \in \mathcal{J}^{\mathrm{f}} \\ k \neq j}} b_{j'k'}^{+}(t)\,\mathrm{d}N_{jk}(t) \Bigg), && B^{\circ,\mathrm{f}}(0) = 0,
\end{aligned}
$$

with $\mathcal{J}^{\mathrm{f}} \ni j \mapsto j' := j - (J+1)$ and $x^{+} := \max\{0, x\}$, and where τ is the time of free policy conversion given by

$$\tau = \inf\{t \in [0,\infty) : Z(t) \in \mathcal{J}^{\mathrm{f}}\}.$$

We have $\tau = 0$ if and only if $z_0 \in \mathcal{J}^{\mathrm{f}}$; in this case, the policy is initially a free policy. Without loss of generality we thus let $\rho(0) = 1$. Furthermore, we suppose there are no sojourn payments in the surrender states, i.e. $b_J \equiv 0$.

It is useful to decompose the predetermined payment stream B° into benefit and premium parts. We add the superscript $\pm$ to denote the benefit and premium part, respectively. Then we have

$$
\begin{aligned}
B^{\circ,-}(t) &= B^{\circ,\mathrm{p},-}(t), \\
B^{\circ,+}(t) &= B^{\circ,\mathrm{p},+}(t) + \rho(\tau) B^{\circ,\mathrm{f}}(t), \\
\mathrm{d}B^{\circ,\mathrm{p},\pm}(t) &= \sum_{j \in \mathcal{J}^{\mathrm{p}}} \mathbb{1}_{(Z(t-)=j)} \bigg(b_j^{\pm}(t)\,\mathrm{d}t + \sum_{\substack{k \in \mathcal{J}^{\mathrm{p}} \\ k \neq j}} b_{jk}^{\pm}(t)\,\mathrm{d}N_{jk}(t) \bigg), \quad B^{\circ,\mathrm{p},\pm}(0) = 0.
\end{aligned}
$$

In the following, we assume the existence of a maximal contract time $n \in (0,\infty)$ in the sense that all sojourn payment rates and transition payments, including those of the unit bonus payment stream, cf. Subsection 3.2.4, are zero for $t > n$.

3.2.3 Valuation of predetermined payments

The life insurance contract is written on the *technical basis*, also called the first order basis, which is at least originally designed to consist of prudent assumptions on financial risk and insurance risk. The technical basis is modeled via another probability measure $\mathrm{P}^\star$ under which the short rate process $r^\star$ is deterministic and suitably regular, while Z is independent of S and Markovian with suitably regular transition rates $\mu^\star$. The assumptions regarding absorption, as illustrated in Figure 3.1, are retained under $\mathrm{P}^\star$. Policyholder behavior is not included on the technical basis, which entails the following constraints on the transition rates, surrender payments, and free policy factor, see Buchardt and Møller (2015) and Buchardt, Møller, and Schmidt (2015):

$$
\begin{aligned}
&\mu_{jk}^\star = \mu_{j'k'}^\star, && j,k \in \mathcal{J}^{\mathrm{f}}, k \neq j, \\
&b_{0J} = \widetilde{V}_0^\star, && \\
&(0,\infty) \ni t \mapsto \rho(t) = \frac{\widetilde{V}_0^\star(t)}{\widetilde{V}_0^{\star,+}(t)}, &&
\end{aligned}
$$

where for $j \in \mathcal{J}^{\mathrm{p}} \backslash \{J\}$ the state-wise technical reserve $\widetilde{V}_j^\star$ of predetermined payments and the corresponding valuation of benefits only $\widetilde{V}^{\star,+}$ are given by

$$
\widetilde{V}_j^\star(t) = \mathrm{E}^\star\bigg[\int_t^n e^{-\int_t^s r^\star(v)\,\mathrm{d}v}\,\mathrm{d}B^\circ(s) \,\bigg|\, Z(t) = j \bigg], \tag{3.2.1}
$$

$$
\widetilde{V}_j^{\star,+}(t) = \mathrm{E}^\star\bigg[\int_t^n e^{-\int_t^s r^\star(v)\,\mathrm{d}v}\,\mathrm{d}B^{\circ,+}(s) \,\bigg|\, Z(t) = j \bigg], \tag{3.2.2}
$$

with $\mathrm{E}^\star$ denoting integration w.r.t. $\mathrm{P}^\star$. It it possible to show that the state-wise technical reserves of predetermined payments satisfy the following differential equations of Thiele type:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{V}_j^\star(t) &= r^\star(t)\widetilde{V}_j^\star(t) - b_j(t) \\
&\quad - \sum_{\substack{k \in \mathcal{J}^{\mathrm{p}} \backslash \{J\} \\ k \neq j}} \big(b_{jk}(t) + \widetilde{V}_k^\star(t) - \widetilde{V}_j^\star(t) \big) \mu_{jk}^\star(t), \qquad \widetilde{V}_j^\star(n) = 0,
\end{aligned} \tag{3.2.3}
$$

for $j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}$. By adding $+$'s as superscripts, one finds an identical system of differential equations concerning the valuation of benefits only.

We are now ready to define the technical reserve of predetermined payments denoted $V^{\star,\circ}$. First, for the purpose of bonus allocation, the definitions of state-wise reserves of predetermined payments are naturally extended from $j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}$ to $j \in \mathcal{J}$ via

$$V_j^{\star,\circ}(t) = \begin{cases} \widetilde{V}_j^{\star}(t) & \text{if } j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}, \\ \rho(\tau)\widetilde{V}_{j'}^{\star,+}(t) & \text{if } j \in \mathcal{J}^{\mathrm{f}} \setminus \{2J+1\}, \\ 0 & \text{if } j \in \{J, 2J+1\}. \end{cases} \tag{3.2.4}$$

The technical reserve of predetermined payments $V^{\star,\circ}$ is then defined according to $V^{\star,\circ}(t) = V_{Z(t)}^{\star,\circ}(t)$. Note that $V_j^{\star,\circ}$ depends on τ in the free policy states, thus being stochastic, while it is deterministic in the premium paying states.

We now turn our attention to valuation under the market basis modeled via P. Here we assume that Z and S are independent and that Z is Markovian with suitably regular transition rates μ. The market reserve V° of predetermined payments is then given by

$$V^{\circ}(t) = \mathrm{E}\left[\int_t^n e^{-\int_t^s r(u)\,\mathrm{d}u}\,\mathrm{d}B^{\circ}(s) \,\middle|\, \mathcal{F}(t)\right] = \int_t^n e^{-\int_t^s f(t,u)\,\mathrm{d}u} A^{\circ}(t,\mathrm{d}s), \tag{3.2.5}$$

with A° the so-called expected accumulated predetermined cash flows given by

$$A^{\circ}(t,s) = \mathrm{E}\big[B^{\circ}(s) - B^{\circ}(t) \,|\, \mathcal{F}^Z(t)\big]. \tag{3.2.6}$$

Denote with p the transition probabilities of Z under P. Following Buchardt and Møller (2015) and Buchardt, Møller, and Schmidt (2015), on $(Z(t) \in \mathcal{J}^{\mathrm{f}})$,

$$A^{\circ}(t,\mathrm{d}s) = \rho(\tau)\sum_{j\in\mathcal{J}^{\mathrm{f}}} p_{Z(t)j}(t,s)\Bigg(b_{j'}^{+}(s) + \sum_{\substack{k\in\mathcal{J}^{\mathrm{f}}\\ k\neq j}} b_{j'k'}^{+}(s)\mu_{jk}(s)\Bigg)\mathrm{d}s, \tag{3.2.7}$$

while on $(Z(t) \in \mathcal{J}^{\mathrm{p}})$,

$$\begin{aligned} A^{\circ}(t,\mathrm{d}s) = &\sum_{j\in\mathcal{J}^{\mathrm{p}}} p_{Z(t)j}(t,s)\Bigg(b_j(s) + \sum_{\substack{k\in\mathcal{J}^{\mathrm{p}}\\ k\neq j}} b_{jk}(s)\mu_{jk}(s)\Bigg)\mathrm{d}s \\ &+ \sum_{j\in\mathcal{J}^{\mathrm{f}}} p_{Z(t)j}^{\rho}(t,s)\Bigg(b_{j'}^{+}(s) + \sum_{\substack{k\in\mathcal{J}^{\mathrm{f}}\\ k\neq j}} b_{j'k'}^{+}(s)\mu_{jk}(s)\Bigg)\mathrm{d}s \end{aligned} \tag{3.2.8}$$

where the so-called ρ-modified transition probabilities p_{jk}^{ρ}, $j \in \mathcal{J}^{\mathrm{p}}$ and $k \in \mathcal{J}$, are defined by $p_{jk}^{\rho}(t,s) = \mathrm{E}[\mathbb{1}_{(Z(s)=k)}\rho(\tau)^{\mathbb{1}_{(\tau\leq s)}} \,|\, Z(t) = j]$ and satisfy for $k \in \mathcal{J}^{\mathrm{f}}$

so-called ρ-modified versions of Kolmogorov's forward differential equations:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s} p^{\rho}_{jk}(t,s) &= \sum_{\substack{\ell \in \mathcal{J}^{\mathrm{f}} \\ \ell \neq k}} p^{\rho}_{j\ell}(t,s)\mu_{\ell k}(s) \\ &\quad + \mathbb{1}_{(k=J+1)} p_{j0}(t,s)\mu_{0k}(s)\rho(s) - p^{\rho}_{jk}(t,s)\mu_{k\bullet}(s), \\ p^{\rho}_{jk}(t,t) &= 0,\end{aligned} \tag{3.2.9}$$

while $p^{\rho}_{jk}(t,s) = p_{jk}(t,s)$ for $k \in \mathcal{J}^{\mathrm{p}}$.

3.2.4 Dividends and bonus

With premiums determined by the principle of equivalence based on the prudent technical basis, the portfolio creates a systematic surplus if everything goes well. This surplus mainly belongs to the insured and is to be paid back in the form of dividends. Following Norberg (1999, 2001), we let $D = \{D(t)\}_{t\geq 0}$ denote the accumulated dividends, and we suppose it only consists of absolutely continuous dividend yields:

$$\mathrm{d}D(t) = \delta(t)\,\mathrm{d}t, \quad D(0) = 0,$$

where $\delta = \{\delta(t)\}_{t\geq 0}$ is suitably regular and $\mathcal{F}$-adapted. In Subsection 3.2.6, we specify the dividend strategy further.

We suppose that the dividends are used as a premium to buy additional benefits on the technical basis corresponding to a so-called unit bonus payment stream $B^{\dagger}$ that only consists of benefits and thus is unaffected by the free policy option. It is given by

$$\mathrm{d}B^{\dagger}(t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{(Z(t-)=j)} \Bigg(b^{\dagger}_j(t)\,\mathrm{d}t + \sum_{\substack{k\in\mathcal{J} \\ k\neq j}} b^{\dagger}_{jk}(t)\,\mathrm{d}N_{jk}(t) \Bigg), \quad B^{\dagger}(0) = 0,$$

where the payment functions in the premium paying states $\mathcal{J}^{\mathrm{p}}$, $b^{\dagger}_j$ and $b^{\dagger}_{jk}$, are suitably regular non-negative deterministic functions with $b^{\dagger}_J \equiv 0$, while

$$\begin{aligned} &b^{\dagger}_j = b^{\dagger}_{j'} \quad \text{and} \quad b^{\dagger}_{jk} = b^{\dagger}_{j'k'}, \qquad j,k \in \mathcal{J}^{\mathrm{f}}, k \neq j, \\ &b^{\dagger}_{0J} = \widetilde{V}^{\star,\dagger}_0, \end{aligned}$$

where for $j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}$ we denote by $\widetilde{V}^{\star,\dagger}_j$ the state-wise technical unit reserves of $B^{\dagger}$ given as (3.2.1) with B° replaced by $B^{\dagger}$. Again, these state-wise technical reserves satisfy differential equations of Thiele type, namely (3.2.3) with added superscripts $\dagger$.

For the purpose of bonus allocation, the state-wise technical unit reserves are naturally extended from $j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}$ to $j \in \mathcal{J}$ via

$$V_j^{\star,\dagger}(t) = \begin{cases} \widetilde{V}_j^{\star,\dagger}(t) & \text{if } j \in \mathcal{J}^{\mathrm{p}} \setminus \{J\}, \\ \widetilde{V}_{j'}^{\star,\dagger}(t) & \text{if } j \in \mathcal{J}^{\mathrm{f}} \setminus \{2J+1\}, \\ 0 & \text{if } j \in \{J, 2J+1\}, \end{cases} \tag{3.2.10}$$

when the technical value of the additional benefits $V^{\star,\dagger}$ reads $V^{\star,\dagger}(t) = V_{Z(t)}^{\star,\dagger}(t)$.

The expected accumulated unit bonus cash flows $A^\dagger$ of $B^\dagger$ on the market basis can be found analogously to A° and read

$$A^\dagger(t, \mathrm{d}s) = a^\dagger(t,s)\,\mathrm{d}s, \tag{3.2.11}$$

$$a^\dagger(t,s) = \sum_{j\in\mathcal{J}} p_{Z(t)j}(t,s)\Bigg(b_j^\dagger(s) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b_{jk}^\dagger(s)\mu_{jk}(s)\Bigg). \tag{3.2.12}$$

The state-wise counterparts are denoted $A_i^\dagger$ and $a_i^\dagger$, $i \in \mathcal{J}$. They satisfy $A_{Z(t)}^\dagger(t, \mathrm{d}s) = a_{Z(t)}^\dagger(t,s)\,\mathrm{d}s = a^\dagger(t,s)\,\mathrm{d}s = A^\dagger(t, \mathrm{d}s)$ by taking the form

$$A_i^\dagger(t, \mathrm{d}s) = a_i^\dagger(t,s)\,\mathrm{d}s, \tag{3.2.13}$$

$$a_i^\dagger(t,s) = \sum_{j\in\mathcal{J}} p_{ij}(t,s)\Bigg(b_j^\dagger(s) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b_{jk}^\dagger(s)\mu_{jk}(s)\Bigg). \tag{3.2.14}$$

Let $Q(t)$ denote the number of additional benefits held at time t. Since δ is used as a premium to buy $B^\dagger$ on the technical basis, we have that

$$\mathrm{d}Q(t) = \frac{\mathrm{d}D(t)}{V_{Z(t)}^{\star,\dagger}(t)} = \frac{\delta(t)}{V_{Z(t)}^{\star,\dagger}(t)}\,\mathrm{d}t, \quad Q(0) = 0. \tag{3.2.15}$$

Imposing this bonus mechanism, the total payment stream consisting of both predetermined payments and bonus payments is given by

$$\mathrm{d}B(t) = \mathrm{d}B^\circ(t) + Q(t)\,\mathrm{d}B^\dagger(t), \quad B(0) = 0. \tag{3.2.16}$$

In this paper, we implicitly think of Q as weakly increasing, although this is not a mathematical requirement. This way of thinking is reflected in the terminology. Along these lines, we define the payment process B^g by

$$B^g(t, \mathrm{d}s) = \mathrm{d}B^\circ(s) + Q(t)\,\mathrm{d}B^\dagger(s), \quad B^g(t,t) = B(t), \tag{3.2.17}$$

and refer to it as the *payments guaranteed at time* $t \geq 0$, while the remaining payments

$$(Q(s) - Q(t))\,\mathrm{d}B^\dagger(s)$$

are referred to as *bonus (payments)*.

In the remainder of the paper, we focus on valuation of the payment stream (3.2.16), in particular the bonus payments. We assume that Q exists and is suitably regular, so that the technical arguments in the remainder of the paper are legitimate. This is an implicit condition that must be checked for any specific model.

3.2.5 Liabilities

Thinking of time zero as now, the present life insurance liabilities of the insurer are described by the market value of the total payment stream B evaluated at time zero:

$$V(0) = \mathrm{E}\left[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v}\,\mathrm{d}B(t)\right].$$

By (3.2.16), this amounts to market valuation of the predetermined payments and bonus payments. Thus $V(0) = V^\circ(0) + V^b(0)$ where $V^\circ(0)$ is given by (3.2.5) and

$$V^b(0) = \mathrm{E}\left[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} Q(t)\,\mathrm{d}B^\dagger(t)\right]. \tag{3.2.18}$$

is the time zero market value of bonus payments.

Remark 3.2.1. By setting $Q(0) = 0$, we think of time zero as the time of initialization of the insurance contract. To determine the market value of bonus payments after initialization of the contract, one could extend the filtration $\mathcal{F}$ to include additional information at time zero and consider a general $\mathcal{F}(0)$-adapted $Q(0)$. This extension is straightforward and achieved by focusing on $Q(\cdot) - Q(0)$ rather than $Q(\cdot)$, and thus the requirement $Q(0) = 0$ is only really made for notational convenience. $\triangledown$

There exists well-established methods to calculate $V^\circ(0)$ explicitly using the expected accumulated cash flows of predetermined payments on the market basis from (3.2.7)–(3.2.8); in particular, this computation does not depend on the dividend strategy δ nor further realizations of the financial market (only the forward rate curve $f(0,\cdot)$ is required). On the contrary, the time zero market value of bonus payments $V^b(0)$ *does* depend on the strategy δ. Due to possibly non-linear path dependencies regarding both the financial and biometric/behavioral scenarios, this implies that classic computational methods via (ρ-modified) Kolmogorov's forward differential equations are not applicable.

The focal point of the paper is to establish methods to calculate the market value of bonus payments $V^b(0)$. We consider an approach that combines simulations of the financial market with more analytical methods for calculations involving the state of the insured. Everything else being equal, this approach should be numerically superior to a pure simulation approach for which one would simulate both the financial market and the state of the insured. To formalize the main idea, we define what we shall term Q-modified transition probabilities (at time 0) for $j \in \mathcal{J}$ by

$$p^{Q}_{z_0 j}(0,t) = \mathrm{E}\big[Q(t)\mathbb{1}_{(Z(t)=j)} \,\big|\, \mathcal{F}^S(t)\big] \tag{3.2.19}$$

for all $t \geq 0$. We immediately have the following result:

Proposition 3.2.2. *Under suitable regularity conditions the time zero market value of the bonus payments is given by*

$$V^b(0) = \mathrm{E}\bigg[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} A^b(0,\mathrm{d}t)\bigg], \tag{3.2.20}$$

$$A^b(0,\mathrm{d}t) = a^b(0,t)\,\mathrm{d}t, \tag{3.2.21}$$

$$a^b(0,t) := \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t)\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big). \tag{3.2.22}$$

Furthermore, if Q is adapted to $\mathcal{F}^S$, then

$$p^{Q}_{z_0 j}(0,t) = Q(t)p_{z_0 j}(0,t), \tag{3.2.23}$$

$$a^b(0,t) = Q(t)a^\dagger(0,t). \tag{3.2.24}$$

Proof. Since $\{Q(t)\}_{t\geq 0}$ is continuous and adapted, it is predictable. Using martingale techniques, we find that

$$V^b(0) = \mathrm{E}\Bigg[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} \sum_{j\in\mathcal{J}} Q(t)\mathbb{1}_{(Z(t-)=j)}\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big)\mathrm{d}t\Bigg].$$

Due to continuity assumptions, we might replace $\mathbb{1}_{(Z(t-)=j)}$ by $\mathbb{1}_{(Z(t)=j)}$. Using the law of iterated expectations and Fubini's theorem, we conclude that

$$\begin{aligned}
&V^b(0)\\
&= \mathrm{E}\Bigg[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} \sum_{j\in\mathcal{J}} \mathrm{E}\big[1_{(Z(t)=j)}Q(t)\,\big|\,\mathcal{F}^S(t)\big]\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big)\mathrm{d}t\Bigg]\\
&= \mathrm{E}\Bigg[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t)\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big)\mathrm{d}t\Bigg]\\
&= \mathrm{E}\bigg[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} a^b(0,t)\,\mathrm{d}t\bigg].
\end{aligned}$$

Furthermore, if Q is $\mathcal{F}^S$-adapted, then the Q-modified transition probabilities satisfy

$$p^{Q}_{z_0 j}(0,t) = \mathrm{E}\big[\mathbb{1}_{(Z(t)=j)}Q(t)\,\big|\,\mathcal{F}^S(t)\big] = Q(t)p_{z_0 j}(0,t),$$

and thus $a^b(0,t) = Q(t)a^\dagger(0,t)$, cf. (3.2.12). $\square$

Since the so-called expected accumulated bonus cash flow $A^b(0,\cdot)$ is $\mathcal{F}^S$-adapted, the result provides a representation of $V^b(0)$ motivating a computational scheme based on simulation of the financial market. For each simulated financial scenario, we should compute $A^b(0,\cdot)$ explicitly in each scenario, which in general requires computation of of $p^Q_{z_0 j}(0,\cdot)$ for all $j \in \mathcal{J}$; this we study in Section 3.3. In the special case where Q is $\mathcal{F}^S$-adapted, it holds that $p^Q_{z_0 j}(0,\cdot) = Q(\cdot)p_{z_0 j}(0,\cdot)$, and the problem simplifies to a direct calculation of Q that does not involve the biometric/behavioral states, and can essentially be solved by a classic computation of the expected accumulated cash flow $A^\dagger(0,\cdot)$ via Kolmogorov's forward differential equations; this is studied in Section 3.4.

As mentioned above, the computation of the expected accumulated bonus cash flow depends on the actual specification of the dividend strategy δ during the course of the contract, and in practice, this strategy is a control variable that depends on what we refer to as the shape of the insurance business. In the following subsection, we formalize the shape of the insurance business and its corresponding controls, which leads to a specification of a class of dividend strategies.

3.2.6 Shape and controls

We now introduce the shape of the insurance business consisting of key quantities on a portfolio level that the insurer needs at future time points to determine the controls, i.e. the dividend strategy and the investment strategy. We only introduce a few key financial indicators, but we believe that our general methodology allows for the implementation of additional shape variables.

To describe the shape of the insurance business, we first consider the liabilities, specifically the technical value and the market value of guaranteed payments on a portfolio level. Recall that the payments $B^g(t,\cdot)$ guaranteed at time $t \geq 0$ take the form (3.2.17). The market value of guaranteed payments V^g is thus given by

$$V^g(t) = \mathrm{E}\left[\int_t^n e^{-\int_t^s r(v)\,\mathrm{d}v} B^g(t,\mathrm{d}s) \,\middle|\, \mathcal{F}(t)\right] = \int_t^n e^{-\int_t^s f(t,v)\,\mathrm{d}v} A^g(t,\,\mathrm{d}s), \quad (3.2.25)$$

with A^g denoting the expected accumulated guaranteed cash flows,

$$A^g(t,\mathrm{d}s) = A^\circ(t,\mathrm{d}s) + Q(t)A^\dagger_{Z(t)}(t,\mathrm{d}s). \quad (3.2.26)$$

Similary, the technical reserve of guaranteed payments is given by

$$V^\star(t) = V^{\star,\circ}(t) + Q(t)V^{\star,\dagger}_{Z(t)}(t). \quad (3.2.27)$$

The so-called *portfolio-wide means* of $V^\star$ and V^g are now obtained by averaging out the unsystematic insurance risk by applying the law of large numbers w.r.t. a collection of independent and comparable insured in the portfolio, see e.g. the discussions in Møller and Steffensen (2007, Chapter 6) and Norberg (1991). The portfolio-wide means take the form

$$\bar{V}^g(t) = \mathrm{E}\big[\, V^g(t) \,|\, \mathcal{F}^S(t)\big] \qquad \text{and} \qquad \bar{V}^\star(t) = \mathrm{E}\big[\, V^\star(t) \,|\, \mathcal{F}^S(t)\big]$$

for $t \geq 0$. The portfolio-wide means represent values of liabilities under the assumption that the insurance portfolio is of such a size that unsystematic insurance risk can be disregarded. It corresponds to what is often referred to as mean-field approximations in the literature. In Subsection 3.3.1, we show how to compute these.

We now turn our attention to the assets. They are described by a portfolio of S which is self-financed by the premium less benefits that the portfolio of insured pays to the insurer. We denote the value process by $U = \{U(t)\}_{t \geq 0}$. We think of this process as the assets for the whole portfolio, but in our presentation the payments involved are only the contributions of a single insured. Since an individual insured pays $-\,\mathrm{d}B(t)$ to the insurer, this contribution to the total payments of the portfolio can be represented by the expected cash flow $-\big(A^\circ(0, \mathrm{d}t) + A^b(0, \mathrm{d}t)\big)$. Thus we let U take the form

$$\mathrm{d}U(t) = \theta(t)\,\mathrm{d}S_0(t) + \eta(t)\,\mathrm{d}S_1(t) - \big(A^\circ(0, \mathrm{d}t) + A^b(0, \mathrm{d}t)\big), \quad U(0) \equiv u_0,$$

where $(\theta, \eta) = (\theta(t), \eta(t))_{t \geq 0}$ is a suitably regular $\mathcal{F}^S$-adapted investment strategy. We think of η as a control variable for the insurer, since the number of units invested into the bank account is determined residually by $\theta(t) = (U(t) - \eta(t)S_1(t))/S_0(t)$. This gives

$$\mathrm{d}U(t) = r(t)(U(t) - \eta(t)S_1(t))\mathrm{d}t + \eta(t)\,\mathrm{d}S_1(t) - \big(A^\circ(0, \mathrm{d}t) + A^b(0, \mathrm{d}t)\big). \tag{3.2.28}$$

In this paper, we only consider a single insured and the portfolio-wide mean reserves represent the contribution of this insured to the shape of the insurance business. To include this observation into the setting, one can consider $Z(0)$ as stochastic with distribution corresponding to the empirical distribution of initial states in the portfolio. The latter can be described by weights w_j with the j'th weigth giving the proportion of insured that are initially in state $j \in \mathcal{J}$. The corresponding portfolio-wide means would in this case read

$$\sum_{j \in \mathcal{J}} w_j \,\mathrm{E}_j\big[\, V^g(t) \,|\, \mathcal{F}^S(t)\big] \qquad \text{and} \qquad \sum_{j \in \mathcal{J}} w_j \,\mathrm{E}_j\big[\, V^\star(t) \,|\, \mathcal{F}^S(t)\big],$$

where E_j corresponds to expectation under the assumption that $Z(0) \equiv j$. Additionally, the insured typically belong to different cohorts implying that e.g. the transition rates and payment processes differ among insured. This is handled in a similar way. Also, the same considerations apply to the payments affecting the value process U. We consider these kinds of extensions from a single insured to a whole portfolio straightforward and do not give them further attention in the remainder of the paper.

Let $S(\cdot \wedge t) = \{S(u)\}_{0 \le u \le t}$. We can now make the concepts of shape and controls precise.

Definition 3.2.3. *The shape of the insurance business $\mathcal{I}$ is the triplet*

$$\mathcal{I} = \left(U(t), \bar{V}^g(t), \bar{V}^\star(t)\right)_{t \ge 0},$$

while the controls are the pair $(\delta(t), \eta(t))_{t \ge 0}$.

Assumption 3.2.4. *We suppose that (δ, η) are chosen such that the setting is well-specified in the sense that Q exists and is suitably regular. Furthermore, we assume that η takes the form*

$$\eta(t) = \eta(t, S(\cdot \wedge t), \mathcal{I}(t)) \tag{3.2.29}$$

for some explicitly computable and suitably regular deterministic mapping η, and we assume that δ takes the form

$$\begin{aligned} \delta(t) = \delta_0 &(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)) \\ &+ \delta_1 (t, S(\cdot \wedge t), Z(t), \mathcal{I}(t))\, \rho(\tau)^{\mathbb{1}_{(\tau \le t)}} \\ &+ \delta_2 (t, S(\cdot \wedge t), Z(t), \mathcal{I}(t))\, Q(t), \end{aligned} \tag{3.2.30}$$

for some suitably regular deterministic mappings δ_0, δ_1 and δ_2 that we are able to compute explicitly.

Remark 3.2.5. In Remark 3.2.1 we discussed the extension to general $Q(0)$ and the idea of focusing on $Q(\cdot) - Q(0)$. By rewriting (3.2.30) in the following manner,

$$\begin{aligned} \delta(t) = \delta_0 &(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)) + \delta_2 (t, S(\cdot \wedge t), Z(t), \mathcal{I}(t))\, Q(0) \\ &+ \delta_1 (t, S(\cdot \wedge t), Z(t), \mathcal{I}(t))\, \rho(\tau)^{\mathbb{1}_{(\tau \le t)}} \\ &+ \delta_2 (t, S(\cdot \wedge t), Z(t), \mathcal{I}(t))\, (Q(t) - Q(0)), \end{aligned}$$

we see how this idea would manifest itself in relation to Assumption 3.2.4. $\triangledown$

In the following, we also use the shorthand notations $t \mapsto \delta_i(t, Z(t))$, $i = 0, 1, 2$, which only highlights $\mathcal{F}^Z$-measurable quantities.

The assumption that the controls depend only on portfolio-wide means rather than actual realizations of the balance sheet and the assets is the key choice of this paper. The risk we hereby account for is only the systematic risk, i.e. the risk that affects all insured.

Note that it is the assumption of δ being dependent on U that makes η a process that affects the payments to the insured, thus justifying it as a control. Note also that we allow δ to depend on Z, τ, and Q, while this is not the case for η. This is since the dividends are allocated to the individual insured while the assets are a portfolio level quantity. The specific affine structure on δ mirrors that of B, cf. (3.2.16). This is important for practical applications, as the following example highlights.

Example 3.2.6 (Second order interest rate)**.** Dividends may arise by accumulating the technical reserve $V^\star$ from (3.2.27) with a second order interest rate r^δ that is determined based on the shape of the insurance business. This is obtained by letting

$$\begin{aligned}\delta(t) &= \big(r^\delta(t) - r^\star(t)\big)V^\star(t),\\ r^\delta(t) &= \Phi(t, S(\cdot \wedge t), \mathcal{I}(t)),\end{aligned}$$

for some explicitly computable and suitably regular mapping Φ. This corresponds to setting

$$\begin{aligned}\delta_0(t,j) &= \big(r^\delta(t) - r^\star(t)\big)1_{(j\in\mathcal{J}^{\mathrm{p}}\setminus\{J\})}\widetilde{V}^\star_j(t),\\ \delta_1(t,j) &= \big(r^\delta(t) - r^\star(t)\big)1_{(j\in\mathcal{J}^{\mathrm{f}}\setminus\{2J+1\})}\widetilde{V}^{\star,+}_{j'}(t)\\ \delta_2(t,j) &= \big(r^\delta(t) - r^\star(t)\big)V^{\star,\dagger}_j(t),\end{aligned}$$

for all $j \in \mathcal{J}$. ∘

The aim of this paper is to develop methods to compute the market value of bonus payments $V^b(0)$. Recall from Proposition 3.2.2 that this can be done via the computation of the expected accumulated bonus cash flow $A^b(0,\cdot)$, which depends on the financial market through Q. To achieve this within the setup of Assumption 3.2.4, we adopt a simulation approach. It follows from (3.2.15) that for a simulated financial scenario, i.e. a realization of the whole path of S, we need the shape of the insurance business $\mathcal{I}(t) = (U(t), \bar{V}^\star(t), \bar{V}^g(t))$ and corresponding controls $(\delta(t), \eta(t))$ for all time points $t \geq 0$. In other words, starting today from time zero, we must *project* the shape of the insurance business and the controls into future time points for each simulated financial scenario.

In the following sections, we formulate our *scenario-based projection models* demonstrating how to project the shape of the insurance business in a specific financial scenario, and how to apply these projections to calculate the expected accumulated bonus cash flow $A^b(0,\cdot)$. Section 3.3 concerns the general case where Q is allowed to be $\mathcal{F}^Z \vee \mathcal{F}^S$-adapted and where we apply (3.2.21)–(3.2.22). In the subsequent Section 3.4 we specialize to Q being state independent (of Z), i.e. $\mathcal{F}^S$-adapted, where we instead can apply the simpler formula (3.2.24).

3.3 Scenario-based projection model

This section contains the main contributions of the paper and provides the foundation for the special case in Section 3.4. In Subsection 3.3.1, we formulate our general scenario-based projection model demonstrating how to project the shape of the insurance business into future time points in a given financial scenario. The projections are then in Subsection 3.3.2 used to calculate the Q-modified transition probabilities $p^{Q}_{z_0 j}(0,\cdot)$ and corresponding expected accumulated bonus cash flow $A^b(0,\cdot)$. Based on this, we present in Subsection 3.3.3 a procedure for the computation of $V^b(0)$ via an application of Proposition 3.2.2.

As noted in Proposition 3.2.2, we are able to simplify calculations of $A^b(0,\cdot)$ to what we coin *state independent* calculations of Q and p if Q is assumed $\mathcal{F}^S$-adapted. This special case leads to a notion of a *state independent scenario-based projection model*, which is studied in more details in Section 3.4.

3.3.1 Projecting the shape

We now turn our attention to projection of the shape of the insurance business. This consists of computation of $\mathcal{I} = (U, \bar{V}^g, \bar{V}^\star)$ for realizations of S, where each realization exactly represents a simulated financial scenario.

The method for computation of U for a realization of S follows immediately from the dynamics of the assets according to (3.2.28). The computational issue reduces to that of computing $p^{Q}_{z_0 j}(0,\cdot)$, cf. (3.2.21)–(3.2.22) and (3.2.28). Thus we focus on the projection of the portfolio-wide means $\bar{V}^g$ and $\bar{V}^\star$.

First, we consider the portfolio-wide mean of the market value of guaranteed payments, $\bar{V}^g$. From (3.2.25), calculation of $\bar{V}^g$ is a matter of calculating the portfolio-wide means $\bar{A}^g$ of the expected accumulated guaranteed cash flows A^g defined by

$$\bar{A}^g(t,s) = \mathrm{E}\big[\, A^g(t,s) \,|\, \mathcal{F}^S(t)\big]$$

for $0 \leq t \leq s < \infty$.

Proposition 3.3.1. *The portfolio-wide means $\bar{A}^g$ of the expected accumulated guaranteed cash flows A^g read*

$$\bar{A}^g(t,\mathrm{d}s) = A^\circ(0,\mathrm{d}s) + \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t) A^\dagger_j(t,\mathrm{d}s)$$

for all $t \geq 0$.

Proof. By (3.2.26), (3.2.19), and due to the assumed independence between Z and S, we immediately find that

$$\begin{aligned}\bar{A}^g(t,s) &= \mathrm{E}\big[\, A^\circ(t,s) \,|\, \mathcal{F}^S(t)\big] + \sum_{j\in\mathcal{J}} \mathrm{E}\Big[\mathbb{1}_{(Z(t)=j)} Q(t) A^\dagger_{Z(t)}(t,s) \,\Big|\, \mathcal{F}^S(t)\Big] \\ &= \mathrm{E}[A^\circ(t,s)] + \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t) A^\dagger_j(t,s).\end{aligned}$$

By (3.2.6) and the iterated law of expectations,

$$\begin{aligned}\mathrm{E}[A^\circ(t,s)] &= \mathrm{E}[B^\circ(s) - B^\circ(t)] \\ &= A^\circ(0,s) - \mathrm{E}[B^\circ(t) - B^\circ(0)].\end{aligned}$$

Since the latter term does not depend on s, we find that

$$\bar{A}^g(t,\,\mathrm{d}s) = A^\circ(0,\,\mathrm{d}s) + \sum_{j\in\mathcal{J}} p^Q_{z_0 j}(0,t)A^\dagger_j(t,\,\mathrm{d}s)$$

as desired. □

Calculation of $\bar{V}^g(t)$ now proceeds by discounting $\bar{A}^g(t,\cdot)$ with the forward rate curve available at time t according to the following expression:

$$\bar{V}^g(t) = \int_t^n e^{-\int_t^s f(t,v)\,\mathrm{d}v}\bar{A}^g(t,\mathrm{d}s). \tag{3.3.1}$$

Consequently, given A° and $A^\dagger$ the computational issue has been reduced to that of computing the Q-modified transition probabilities $p^Q_{z_0 j}(0,\cdot)$.

Next we consider the portfolio-wide mean of the technical reserve of guaranteed payments, $\bar{V}^\star$. We could follow the same approach above and calculate the technical reserves via expected (accumulated) cash flows, however, since the technical interest rate is deterministic, a range of technical reserves, including $V^{\star,\dagger}$, $\widetilde{V}^\star$, and $\widetilde{V}^{\star,+}$, can be computed more efficiently by solving the differential equations of Thiele type derived from (3.2.3), cf. Subsection 3.2.3 and Subsection 3.2.4.

Denote by $\bar{V}^{\star,\circ}$ the portfolio-wide mean technical reserves of predetermined payments given by

$$\bar{V}^{\star,\circ}(t) = \mathrm{E}\big[\,V^{\star,\circ}(t)\,|\,\mathcal{F}^S(t)\big]$$

for $t \geq 0$. Since Z and S are assumed independent, we could replace the conditional expectation by an ordinary expectation.

Proposition 3.3.2. *The portfolio-wide mean technical reserve of guaranteed payments reads*

$$\bar{V}^\star(t) = \bar{V}^{\star,\circ}(t) + \sum_{j\in\mathcal{J}} p^Q_{z_0 j}(0,t)V^{\star,\dagger}_j(t),$$

while the portfolio-wide mean technical reserve of predetermined payments reads

$$\bar{V}^{\star,\circ}(t) = \sum_{\substack{j\in\mathcal{J}^{\mathrm{p}}\\ j\neq J}} p_{z_0 j}(0,t)\widetilde{V}^\star_j(t) + \sum_{\substack{j\in\mathcal{J}^{\mathrm{f}}\\ j\neq 2J+1}} p^\rho_{z_0 j}(0,t)\widetilde{V}^{\star,+}_{j'}(t). \tag{3.3.2}$$

Proof. By (3.2.27) and (3.2.19), direct calculations yield

$$\begin{aligned}\bar{V}^{\star}(t) &= \mathrm{E}\big[V^{\star,\circ}(t)\,|\,\mathcal{F}^S(t)\big] + \sum_{j\in\mathcal{J}} \mathrm{E}\Big[\mathbb{1}_{(Z(t)=j)}Q(t)V^{\star,\dagger}_{Z(t)}(t)\,\Big|\,\mathcal{F}^S(t)\Big] \\ &= \bar{V}^{\star,\circ}(t) + \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t)V^{\star,\dagger}_j(t).\end{aligned}$$

To obtain (3.3.2), we split $V^{\star,\circ}$ according to the events of $Z(t)$ being in $\mathcal{J}^{\mathrm{p}}\setminus\{J\}$, $\mathcal{J}^{\mathrm{f}}\setminus\{2J+1\}$, and $\{J, 2J+1\}$. According to (3.2.4), we then have

$$\begin{aligned}\bar{V}^{\star,\circ}(t) &= \mathrm{E}\Big[\mathbb{1}_{(Z(t)\in\mathcal{J}^{\mathrm{p}}\setminus\{J\})}\widetilde{V}^{\star}_{Z(t)}(t) + \mathbb{1}_{(Z(t)\in\mathcal{J}^{\mathrm{f}}\setminus\{2J+1\})}\rho(\tau)\widetilde{V}^{\star,+}_{Z(t)'}(t)\,\Big|\,\mathcal{F}^S(t)\Big] \\ &= \mathrm{E}\Bigg[\sum_{\substack{j\in\mathcal{J}^{\mathrm{p}}\\ j\neq J}} \mathbb{1}_{(Z(t)=j)}\widetilde{V}^{\star}_j(t) + \sum_{\substack{j\in\mathcal{J}^{\mathrm{f}}\\ j\neq 2J+1}} \mathbb{1}_{(Z(t)=j)}\rho(\tau)\widetilde{V}^{\star,+}_{j'}(t)\,\Bigg|\,\mathcal{F}^S(t)\Bigg] \\ &= \sum_{\substack{j\in\mathcal{J}^{\mathrm{p}}\\ j\neq J}} p_{z_0 j}(0,t)\widetilde{V}^{\star}_j(t) + \sum_{\substack{j\in\mathcal{J}^{\mathrm{f}}\\ j\neq 2J+1}} p^{\rho}_{z_0 j}(0,t)\widetilde{V}^{\star,+}_{j'}(t),\end{aligned}$$

as desired. □

As already mentioned, the technical reserves $V^{\star,\dagger}$, $\widetilde{V}^{\star}$, and $\widetilde{V}^{\star,+}$ can be computed efficiently using differential equations of Thiele type, while the ρ-modified transition probabilities are simply computed according to (3.2.9). Thus Proposition 3.3.2 reduces the computational complexity to that of computing Q-modified transition probabilities $p^{Q}_{z_0 j}(0,\cdot)$. This computation is studied in details in the next subsection.

3.3.2 Q-modified transition probabilities

We are now ready to present a system of differential equations for the Q-modified transition probabilities $p^{Q}_{z_0 j}(0,\cdot)$; here $p^{\rho}_{z_0 j}(0,\cdot) := p_{z_0 j}(0,\cdot)$ for $z_0 \in \mathcal{J}^{\mathrm{f}}$, which is in accordance with $\tau = 0$ for $z_0 \in \mathcal{J}^{\mathrm{f}}$ and the assumption $\rho(0) = 1$.

Theorem 3.3.3. *The Q-modified transition probabilities $p^{Q}_{z_0 j}(0,\cdot)$ satisfy for $j \in \mathcal{J}$ the differential equations*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p^{Q}_{z_0 j}(0,t) &= \frac{p_{z_0 j}(0,t)\delta_0(t,j) + p^{\rho}_{z_0 j}(0,t)\delta_1(t,j) + p^{Q}_{z_0 j}(0,t)\delta_2(t,j)}{V^{\star,\dagger}_j(t)} \\ &\quad - p^{Q}_{z_0 j}(0,t)\mu_{j\bullet}(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} p^{Q}_{z_0 k}(0,t)\mu_{kj}(t), \qquad (3.3.3)\\ p^{Q}_{z_0 j}(0,0) &= 0.\end{aligned}$$

Proof. The boundary conditions follows by the assumption that $Q(0) = 0$. Referring to (3.2.19) and (3.2.15), we have

$$p^{Q}_{z_0 j}(0,t) = \mathrm{E}\big[\mathbb{1}_{(Z(t)=j)}Q(t)\,\big|\,\mathcal{F}^S(t)\big] = \mathrm{E}\Bigg[\mathbb{1}_{(Z(t)=j)}\int_0^t \frac{\delta(u)}{V^{\star,\dagger}_{Z(u)}(u)}\,\mathrm{d}u\,\Bigg|\,\mathcal{F}^S(t)\Bigg]$$

with

$$\delta(t) = \delta_0(t, Z(t)) + \delta_1(t, Z(t))\rho(\tau)^{\mathbb{1}_{(\tau \leq t)}} + \delta_2(t, Z(t))Q(t).$$

Note that for $0 \leq u \leq t$ and $k \in \mathcal{J}$,

$$\begin{aligned}
\mathrm{E}\left[\mathbb{1}_{(Z(u)=k)} \frac{p^Q_{z_0 k}(0,u)}{p_{z_0 k}(0,u)} \,\middle|\, \mathcal{F}^S(t)\right] &= \mathrm{E}\big[\mathbb{1}_{(Z(u)=k)} Q(u) \,\big|\, \mathcal{F}^S(t)\big], \\
\mathrm{E}\left[\mathbb{1}_{(Z(u)=k)} \frac{p^\rho_{z_0 k}(0,u)}{p_{z_0 k}(0,u)}\right] &= \mathrm{E}\big[\mathbb{1}_{(Z(u)=k)} \rho(\tau)^{\mathbb{1}_{(\tau \leq u)}}\big].
\end{aligned}$$

Thus by Markovianity of Z and independence between Z and S,

$$p^Q_{z_0 j}(0,t) = \mathrm{E}\left[\mathbb{1}_{(Z(t)=j)} \int_0^t \sum_{k \in \mathcal{J}} \mathbb{1}_{(Z(u)=k)} b^Q_k(u)\,\mathrm{d}u \,\middle|\, \mathcal{F}^S(t)\right] \tag{3.3.4}$$

with b^Q_k, $k \in \mathcal{J}$, given by

$$b^Q_k(u) = \frac{\delta_0(u,k) + \delta_1(u,k)\frac{p^\rho_{z_0 k}(0,u)}{p_{z_0 k}(0,u)} + \delta_2(u,k)\frac{p^Q_{z_0 k}(0,u)}{p_{z_0 k}(0,u)}}{V^{\star,\dagger}_k(u)} \tag{3.3.5}$$

for all $u \geq 0$. The assumption of independence between Z and S, Markovianity of Z, and Fubini's theorem finally yield

$$p^Q_{z_0 j}(0,t) = \int_0^t \sum_{k \in \mathcal{J}} p_{z_0 k}(0,u) p_{kj}(u,t) b^Q_k(u)\,\mathrm{d}u. \tag{3.3.6}$$

The statement of the theorem is now established by differentiation as follows. Leibniz' integration rule gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p^Q_{z_0 j}(0,t) &= \sum_{k \in \mathcal{J}} 1_{(k=j)} p_{z_0 k}(0,t) b^Q_k(t) + \int_0^t \sum_{k \in \mathcal{J}} p_{z_0 k}(0,u)\left(\frac{\mathrm{d}}{\mathrm{d}t} p_{kj}(u,t)\right) b^Q_k(u)\,\mathrm{d}u \\
&= \frac{\delta_0(t,j) p_{z_0 j}(0,t) + \delta_1(t,j) p^\rho_{z_0 j}(0,t) + \delta_2(t,j) p^Q_{z_0 j}(0,t)}{V^{\star,\dagger}_j(t)} \\
&\quad + \int_0^t \sum_{k \in \mathcal{J}} p_{z_0 k}(0,u)\left(\frac{\mathrm{d}}{\mathrm{d}t} p_{kj}(u,t)\right) b^Q_k(u)\,\mathrm{d}u.
\end{aligned}$$

Applying Kolmogorov's forward differential equations and (3.3.6) to the last line of the equation we find that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p^Q_{z_0 j}(0,t) &= \frac{\delta_0(t,j) p_{z_0 j}(0,t) + \delta_1(t,j) p^\rho_{z_0 j}(0,t) + \delta_2(t,j) p^Q_{z_0 j}(0,t)}{V^{\star,\dagger}_j(t)} \\
&\quad - p^Q_{z_0 j}(0,t)\mu_{j\bullet}(t) + \sum_{\substack{\ell \in \mathcal{J} \\ \ell \neq j}} p^Q_{z_0 \ell}(0,t)\mu_{\ell j}(t)
\end{aligned}$$

as desired. $\square$

Remark 3.3.4. There exists a clear link between Q-modified transition probabilities and so-called state-wise retrospective reserves. Referring to (3.3.4) and (3.3.5), we see that for a fixed financial scenario,

$$W_j(\cdot) := \frac{p^Q_{z_0 j}(0,\cdot)}{p_{z_0 j}(0,\cdot)}$$

corresponds to the state-wise retrospective reserve of Norberg (1991) (in the presence of information $\mathcal{G}(t) = \mathcal{F}^S(t) \vee \sigma(Z(t))$, cf. Norberg, 1991, Subsection 5.B) with payments

$$-\sum_{j\in\mathcal{J}} \mathbb{1}_{(Z(t)=j)} b^Q_j(t)\,\mathrm{d}t$$

and interest rate zero. Contrary to the primary setup of Norberg (1991), the payments considered here are functions of the state-wise retrospective reserves $W_j(\cdot)$. ▽

The system of differential equations for $p^Q_{z_0 j}(0,\cdot)$ from Theorem 3.3.3 involves the shape of the insurance business $\mathcal{I}$ through the mappings δ_0, δ_1, and δ_2. Together with the results of the previous subsection, Theorem 3.3.3 allows us formulate a procedure for the calculation of $V^b(0)$. The procedure is presented in the next subsection.

3.3.3 Numerical procedure

Based on the results of the previous subsections, we demonstrate a procedure for the scenario-based projection model. In what follows, we suppose we are given mappings (δ, η) serving as controls. They are assumed to satisfy Assumption 3.2.4.

Besides the financial scenarios, the input consists of the following quantities which can be precalculated independently of the financial scenarios:

(1) The expected accumulated cash flow of predetermined payments $A^\circ(0,s)$ for $s \geq 0$ as in (3.2.8).

(2) The portfolio-wide mean technical reserve of predetermined payments $\bar{V}^{\star,\circ}(t)$ for all $t \geq 0$ calculated via (3.3.2).

(3) For each $t \geq 0$, state-wise expected accumulated unit bonus cash flows $A^\dagger_j(t,s)$ for all $s \geq t$ and $j \in \mathcal{J}$ as in (3.2.13)–(3.2.14).

(4) State-wise technical unit reserves $V^{\star,\dagger}_j(t)$ for all $t \geq 0$ and $j \in \mathcal{J}$ as in (3.2.10).

(5) Transition probabilities $p_{z_0 j}(0,t)$ for all $t \geq 0$ and $j \in \mathcal{J}$.

As discussed previously, this input can be calculated using classic methods for solving differential equations of Thiele type as well as (ρ-modified) Kolmogorov forward differential equations.

The financial scenarios are N realizations $\{S^k(t)\}_{t\geq 0}$, $k = 1, \ldots, N$, of $\{S(t)\}_{t\geq 0}$ with corresponding short rate r^k and forward rate curves f^k. We consider them as output of an economic scenario generator.

The procedure essentially consists of computing $p^Q_{z_0 j}(0, \cdot)$, $j \in \mathcal{J}$, and $U(\cdot)$ in each financial scenario by solving a system of (stochastic) differential equations. The involved part is to evaluate the differentials. The procedure looks as follows. For each financial scenario $k = 1, \ldots, N$:

- Initialize with $p^{Q,k}_{z_0 j}(0, 0) = 0$ for all $j \in \mathcal{J}$ and $U^k(0) = u_0$.
- Apply a numerical algorithm to solve the coupled (stochastic) differential equation systems for $p^{Q,k}_{z_0 j}(0, \cdot)$, $j \in \mathcal{J}$, and $U^k(\cdot)$ from Theorem 3.3.3 and (3.2.28), respectively.
  - Evaluating the differentials at time t involves the mappings $(\delta_0, \delta_1, \delta_2, \eta)$ from (3.2.29)–(3.2.30). By inspection of the differentials and these mappings, we see that we require the shape of the insurance business

    $$\mathcal{I}^k(t) = \big(U^k(t), \bar{V}^{g,k}(t), \bar{V}^{\star,k}(t)\big),$$

    the expected bonus cash flow $a^{b,k}(0, t)$, as well as the input. Computation of $\bar{V}^{g,k}(t)$, $\bar{V}^{\star,k}(t)$, and $a^{b,k}(0, t)$ is achieved via Proposition 3.3.1, Proposition 3.3.2, and (3.2.22).
- We emphasize that as part of evaluating the differentials we computed the expected bonus cash flow $a^{b,k}(0, \cdot)$.

The procedure completes by computing the market value of bonus payments $V^b(0)$ via

$$V^b(0) \approx \frac{1}{N} \sum_{k=1}^{N} \int_0^n e^{-\int_0^t r^k(v)\,\mathrm{d}v} a^{b,k}(0, t)\,\mathrm{d}t$$

using an algorithm for numerical integration.

Note that we require the input (3), which are the state-wise expected accumulated unit bonus cash flows $A^\dagger_j(\cdot, \cdot)$ evaluated on the two-dimensional time grid $\{(t, s) \in [0, \infty)^2 : t \leq s\}$. To precompute this input, one must solve Kolmogorov's forward differential equations many times, once for every $t \geq 0$ and $j \in \mathcal{J}$. This significantly impacts the numerical efficiency of the procedure. Furthermore, the algorithm itself depends on the market basis for the specific insured through the transition rates μ.

In practice, where the algorithm must be executed for many insured, one must view the specific transition rates for a single insured as input.

In the following section, we present the simpler state independent scenario-based projection model, where we require that the dividend strategy be specified (or approximated) such that Q is $\mathcal{F}^S$-adapted. By presenting a numerical procedure for the model, we show how this requirement on the dividend strategies leads to a numerical speedup.

3.4 State independent scenario-based projection model

This section concerns the formulation of the state independent scenario-based projection model. The model is a special case of the projection model from Section 3.3 which relies on ensuring Q to be an $\mathcal{F}^S$-adapted process such that the simplified case of Proposition 3.2.2 applies. In Subsection 3.4.1, we provide sufficient conditions on δ such that Q is $\mathcal{F}^S$-adapted. Next, Subsection 3.4.2 revisits the projection of the shape under this simplification. Finally, in Subsection 3.4.3 we present a procedure for the computation of the market value of bonus payments in the state independent projection model.

3.4.1 Class of dividend strategies

Recall from (3.2.15) and (3.2.30) that Q is the solution to the differential/integral equation

$$\mathrm{d}Q(t) = \frac{\delta_0(t, Z(t)) + \delta_1(t, Z(t))\rho(\tau)^{\mathbb{1}_{(\tau \leq t)}} + \delta_2(t, Z(t))Q(t)}{V^{\star,\dagger}_{Z(t)}(t)}\,\mathrm{d}t, \quad Q(0) = 0.$$

To ensure that Q is an $\mathcal{F}^S$-adapted process, it suffices to require that δ_0, δ_1 and δ_2 are on the form

$$\delta_i(t, Z(t)) = \widetilde{\delta}_i(t)V^{\star,\dagger}_{Z(t)}(t), \qquad i = 0, 2, \tag{3.4.1}$$

$$\delta_1(t, Z(t)) = 0, \tag{3.4.2}$$

where we have used the shorthand notation $\widetilde{\delta}_i(t) = \widetilde{\delta}_i\left(t, S(\cdot \wedge t), \mathcal{I}(t)\right)$ for suitably regular deterministic mappings $\widetilde{\delta}_i$, $i = 0, 2$. This is a consequence of the following observation. When (3.4.1)–(3.4.2) hold, then simply

$$\mathrm{d}Q(t) = \big(\widetilde{\delta}_0(t) + \widetilde{\delta}_2(t)Q(t)\big)\mathrm{d}t, \quad Q(0) = 0. \tag{3.4.3}$$

This implies $p^Q_{z_0 j}(0, t) = Q(t)p_{z_0 j}(0, t)$, cf. (3.2.23).

Remark 3.4.1. Since the class of dividend strategies presented here builds on Assumption 3.2.4, affinity in Q is more or less implicitly assumed. The simplifications we obtain in the following Subsections 3.4.2–3.4.3 build on Q being $\mathcal{F}^S$-adapted rather than the dividend strategy being affine in Q. The results are therefore trivially extendable to dividend strategies that are non-affine in the number of additional benefits held. $\triangledown$

3.4.2 Projecting the shape revisited

For the portfolio-wide means $\bar{A}^g$ we observe a simplification in the part that concerns future bonus payments similar to what we previously saw concerning the predetermined payments:

Corollary 3.4.2. *Assume that the dividend strategy δ is on the form* (3.4.1)–(3.4.2). *The portfolio-wide means $\bar{A}^g$ of the expected accumulated guaranteed cash flows A^g then read*

$$\bar{A}^g(t, \mathrm{d}s) = A^\circ(0, \mathrm{d}s) + Q(t)A^\dagger(0, \mathrm{d}s).$$

Proof. From Proposition 3.3.1 and its proof, we have

$$\bar{A}^g(t, s) = A^\circ(0, s) - E[B^\circ(t) - B^\circ(0)] + \mathrm{E}\big[Q(t)A^\dagger(t, s) \,\big|\, \mathcal{F}^S(t)\big] .$$

Since by assumption Q is $\mathcal{F}^S$-adapted and Z and S are independent, referring to (3.2.5) with superscript $\circ$ replaced by $\dagger$ and applying the law of iterated expectations yields

$$\begin{aligned}\mathrm{E}\big[Q(t)A^\dagger(t, s) \,\big|\, \mathcal{F}^S(t)\big] &= Q(t)\,\mathrm{E}\big[B^\dagger(s) - B^\dagger(t)\big] \\ &= Q(t)A^\dagger(0, s) - Q(t)\,\mathrm{E}\big[B^\dagger(t) - B^\dagger(0)\big]\end{aligned}$$

Consequently,

$$\bar{A}^g(t,\, \mathrm{d}s) = A^\circ(0,\, \mathrm{d}s) + Q(t)A^\dagger(0,\, \mathrm{d}s)$$

as desired. $\square$

For the technical reserve, the result is similar. Before we present the result, let the portfolio-wide mean technical unit bonus reserve $\bar{V}^{\star,\dagger}$ be given by

$$\bar{V}^{\star,\dagger}(t) = \mathrm{E}\Big[V^{\star,\dagger}_{Z(t)}(t) \,\Big|\, \mathcal{F}^S(t)\Big]$$

for $t \geq 0$. Since Z and S are assumed independent, we could replace the conditional expectation by an ordinary expectation. It is then a trivial observation that

$$\bar{V}^{\star,\dagger}(t) = \sum_{j \in \mathcal{J}} p_{z_0 j}(0, t)V^{\star,\dagger}_j(t). \tag{3.4.4}$$

Corollary 3.4.3. *Assume that the dividend strategy δ is on the form* (3.4.1)–(3.4.2). *The portfolio-wide mean technical reserve of guaranteed payments then reads*

$$\bar{V}^\star(t) = \bar{V}^{\star,\circ}(t) + Q(t)\bar{V}^{\star,\dagger}(t).$$

Proof. Since by assumption, Q is $\mathcal{F}^S$-adapted and Z and S are independent, the result follows immediately from (3.2.23), Proposition 3.3.2, and (3.4.4). □

The following example is a continuation of Example 3.2.6 regarding the accumulation of the technical reserve with a second order interest rate.

Example 3.4.4 (Second order interest rate continued)**.** The dividend strategy from Example 3.2.6 regarding accumulation of the technical reserve $V^\star$ with a second order interest rate r^δ does not satisfy the requirements on δ from (3.4.1)–(3.4.2). Instead, the strategy

$$\delta(t) = \big(r^\delta(t) - r^\star(t)\big)\frac{\bar{V}^\star(t)}{\bar{V}^{\star,\dagger}(t)} V^{\star,\dagger}_{Z(t)}(t), \tag{3.4.5}$$

satisfies (3.4.1)–(3.4.2) with

$$\widetilde{\delta}_0(t) = (r^\delta(t) - r^\star(t))\frac{\bar{V}^{\star,\circ}(t)}{\bar{V}^{\star,\dagger}(t)} \quad \text{and} \quad \widetilde{\delta}_2(t) = (r^\delta(t) - r^\star(t)).$$

One may think of this strategy as an accumulation of the portfolio-wide mean technical reserve $\bar{V}^\star$ with r^δ instead, since by (3.4.3),

$$\bar{V}^{\star,\dagger}(t)\,\mathrm{d}Q(t) = \big(r^\delta(t) - r^\star(t)\big)\bar{V}^\star(t)\,\mathrm{d}t.$$

By multiplying the strategy (3.4.5) with

$$\frac{V^\star(t)}{\bar{V}^\star(t)} \quad \text{and} \quad \frac{\bar{V}^{\star,\dagger}(t)}{V^{\star,\dagger}_{Z(t)}(t)}$$

one arrives at strategy of Example 3.2.6. If the two ratios are close to one, the strategy (3.4.5) approximates the strategy of Example 3.2.6. Note that

$$\mathrm{E}\big[V^\star(t)/\bar{V}^\star(t)\,\big|\,\mathcal{F}^S(t)\big] = 1,$$

i.e. the portfolio-wide mean of the first ratio is equal to one. For the latter ratio, this is not necessarily the case since it is non-linear in $V^{\star,\dagger}_{Z(t)}(t)$. ∘

3.4.3 Numerical procedure

Based on the results of the previous subsections, we demonstrate a procedure for the state independent scenario-based projection model. In what follows, we suppose we are given mappings (δ, η) serving as controls. They are assumed to satisfy Assumption 3.2.4 with δ on the form (3.4.1)–(3.4.2).

Besides the financial scenarios, the input consists of the following quantities which can be precalculated independently of the financial scenarios:

(1) The expected accumulated cash flow of predetermined payments $A^{\circ}(0,s)$ for all $s \geq 0$ as in (3.2.8).

(2) The portfolio-wide mean technical reserve of predetermined payments $\bar{V}^{\star,\circ}(t)$ for all $t \geq 0$ calculated via (3.3.2).

(3) The expected unit bonus cash flow $a^{\dagger}(0,s)$ for all $s \geq 0$ as in (3.2.12).

(4) The portfolio-wide mean technical unit bonus reserve $\bar{V}^{\star,\dagger}(t)$ for all $t \geq 0$ calculated via (3.4.4)

As discussed previously, this input can be calculated using classic methods for solving differential equations of Thiele type as well as (ρ-modified) Kolmogorov forward differential equations.

The financial scenarios are N realizations $\{S^k(t)\}_{t\geq 0}$, $k = 1,\ldots,N$, of $\{S(t)\}_{t\geq 0}$ with corresponding short rate r^k and forward rate curves f^k. We consider them as output of an economic scenario generator.

The procedure essentially consists of computing $Q(\cdot)$ and $U(\cdot)$ in each financial scenario by solving a system of (stochastic) differential equations. The involved part is to evaluate the differentials. The procedure looks as follows. For each financial scenario $k = 1,\ldots,N$:

- Initialize with $Q^k(0) = 0$ and $U^k(0) = u_0$.
- Apply a numerical algorithm to solve the coupled (stochastic) differential equation systems for $Q^k(\cdot)$ and $U^k(\cdot)$ from (3.4.3) and (3.2.28), respectively.
  - Evaluating the differentials at time t involves the mappings $(\widetilde{\delta}_0, \widetilde{\delta}_2, \eta)$ from (3.2.29) and (3.4.1). By inspection of the differentials and these mappings, we see that we require the shape of the insurance business

    $$\mathcal{I}^k(t) = \big(U^k(t), \bar{V}^{g,k}(t), \bar{V}^{\star,k}(t)\big),$$

    the expected bonus cash flow $a^{b,k}(0,t) = Q^k(t)a^{\dagger}(0,t)$, cf. (3.2.24), as well as the input. Computation of $\bar{V}^{g,k}(t)$ and $\bar{V}^{\star,k}(t)$ is achieved via Corollary 3.4.2 and Corollary 3.4.3.
- We emphasize that as part of evaluating the differentials we computed the expected bonus cash flow $a^{b,k}(0,\cdot)$.

The procedure completes by computing the market value of bonus payments $V^b(0)$ via

$$V^b(0) \approx \frac{1}{N}\sum_{k=1}^{N}\int_0^n e^{-\int_0^t r^k(v)\,\mathrm{d}v} a^{b,k}(0,t)\,\mathrm{d}t$$

using an algorithm for numerical integration.

Note that in comparison with the procedure of Subsection 3.3.3, the expected unit bonus cash flows $a_j^\dagger(t,\cdot)$, $j \in \mathcal{J}$, have only to be precomputed for $j = z_0$ and $t = 0$. This leads to a speedup. Additionally, the procedure itself does not depend on the market basis for the specific insured (except potentially through the mappings $\widetilde{\delta}_0$, $\widetilde{\delta}_2$, and η). These are the primary practical advantages that are gained by strengthening the requirements on the dividend strategy to (3.4.1)–(3.4.2).

3.5 Outlook

In this section, we compare our methodology and results with recent advances in the literature and discuss possible extension in demand by practitioners. Subsection 3.5.1 contains comparisons with Bruhn and Lollike (2020), Falden and Nyegaard (2020), and Jensen and Schomacker (2015), while the inclusion of both duration effects (so-called semi-Markovianity) and the bonus scheme *consolidation* is the focal point of Subsection 3.5.2.

3.5.1 Comparison with recent advances in the literature

In Bruhn and Lollike (2020) and the follow-up paper Falden and Nyegaard (2020), where the methods and results of the former are generalized to allow for surrender and free policy conversion, primary attention is given to the derivation of differential equations for quantities such as

$$\mathrm{E}\big[\mathbb{1}_{(Z(t)=j)} V^\star(t) \,\big|\, \mathcal{F}^S(t)\big].$$

Since $V^\star = V^{\star,\circ} + Q \cdot V^{\star,\dagger}$, we find that $t \mapsto \mathbb{1}_{(Z(t)=j)} V^\star(t)$ is an affine function of $t \mapsto \mathbb{1}_{(Z(t)=j)} Q(t)$. Thus disregarding free policy conversion, we see a direct link between the differential equations derived in Bruhn and Lollike (2020) and Falden and Nyegaard (2020) and those of Theorem 3.3.3. For these results suitable affinity of the dividend strategy is a key assumption.

The inclusion of the policyholder option of free policy conversion adds an additional layer of complexity. We assumed the unit bonus payment stream $B^\dagger$ to be unaffected by the free policy option, which leads to the total payment stream given by (3.2.16). No such assumption is made in Falden and Nyegaard (2020), which leads to more involved payment streams, although by setting $B^\dagger = B^{\circ,+}$, our payment stream equals that of Falden and Nyegaard (2020, Subsection 4.2, cf. (11)–(12)).

We consider some key concepts and provide practical insights that are not within the scope of Bruhn and Lollike (2020) and Falden and Nyegaard (2020). We explicitly include financial risk, which serves as a good starting point for the extension to doubly stochastic models with dependence between the financial market and the stochastic transition rates. Moreover, we identify and discuss the theoretical and practical challenges arising from the fact that the dividend strategy depends on the shape of the insurance business. Furthermore, we provide ready-to-implement numerical schemes for the computation of the market value of bonus payments. Finally, we discuss potential simplifications arising when the number of additional benefits is (approximated to be) $\mathcal{F}^S$-adapted – the state independent case, which might be of particular interest to practitioners.

The projection model described in Jensen and Schomacker (2015, Section 4) appears to be conceptually very close to exactly our state independent model. As an example, additional benefits are in Jensen and Schomacker (2015, see p. 196) bought according to the portfolio-wide mean $\bar{V}^{\star,\dagger}$ of the technical reserve rather than the actual technical reserve $V^{\star,\dagger}_{Z(\cdot)}$; this is exactly in the spirit of our Example 3.4.4. Consequently, we believe that our presentation among other things serves to formalize and generalize the pragmatic approach found in Jensen and Schomacker (2015) and, correspondingly, aims at bridging the gap between the methods and results found in Bruhn and Lollike (2020) and Falden and Nyegaard (2020) and Jensen and Schomacker (2015).

3.5.2 Extensions

In both theory and practice, the generalization to so-called semi-Markovian models introducing duration dependence in the transition rates and payments is popular and impactful, cf. Hoem (1972), Helwich (2008), Christiansen (2012), and Buchardt, Møller, and Schmidt (2015). We believe that the methods we use here can easily be adapted to semi-Markovian models.

The increase in numerical speed from the general case to the state independent case is increasing in the complexity of the intertemporal dependence structure, which can be seen as follows. Referring to Subsection 3.3.3 and Subsection 3.4.3, the general projection model requires as input the expected unit bonus cash flows evaluated on a two-dimensional time grid, while evaluation on a one-dimensional time grid suffices for the state independent model. When including duration effects, the complexity increases, which ought to entail a four-dimensional time/duration grid for the expected unit bonus cash flows in general projections and a two-dimensional time/duration grid in state independent projections. The gain in numerical speed by assuming the state independent special case is thus far greater in the semi-Markovian model compared to the Markovian model.

In Denmark, the bonus scheme known simply as *consolidation* (in Danish: *styrkelse*) sees widespread use in practice, cf. Jensen and Schomacker (2015, Subsection 4.1). Consolidation involves two technical bases: a low (more prudent) basis and a high (less prudent) basis. At the onset of the contract, the predetermined payments, i.e. the payments guaranteed at time zero, satisfy an equivalence principle for which some payments are valuated on the high technical basis and the remaining payments are valuated on the low technical basis. Dividends are then used to shift these payments from the high to the low basis while upholding the relevant equivalence principle. Typically consolidation is combined with the bonus scheme additional benefits in the following manner. When all predetermined payments have been shifted to the low technical basis, future dividends are used to buy additional benefits. This ruins a key affinity assumption, which increases the complexity significantly. In particular, an extension of Theorem 3.3.3 appears to require more sophisticated methods. In the state independent case, the assumption of affinity is not required, cf. Remark 3.4.1. Consequently, we believe that it is straightforward to extend the state independent projection model to include consolidation in combination with additional benefits.

Acknowledgments and declarations of interest

Our work was at first motivated by unpublished notes by Thomas Møller. We thank Thomas Møller for fruitful discussions and for sharing his insights with us.

Christian Furrer's research is partly funded by the Innovation Fund Denmark (IFD) under File No. 7038-00007B. The authors declare no competing interests.

Chapter 4

Representation of scaled expected insurance cash flows via change of measure techniques

Abstract

We consider general life insurance payment processes and study the expected accumulated cash flows that arise when modifying the payments by scaling factors depending on the time of occurrence of specific events. Such modified payment processes arise naturally in the context of incidental policyholder behavior. We associate to the modifications new probability measures which allows for standard representation of the expected accumulated cash flows. The measures are characterized in terms of the original measure and the scaling factors.

Keywords: Life insurance; Incidental policyholder behavior; Jump processes; Föllmer measures

4.1 Introduction

There has recently been an increasing interest in the representation and efficient computation of expected accumulated life insurance cash flows in the presence of free policy behavior and stochastic retirement, see Henriksen et al. (2014), Buchardt, Møller, and Schmidt (2015), Buchardt and Møller (2015), Gad and Nielsen (2016), and Asmussen and Steffensen (2020). In these investigations, the jump process governing the state of the insured is assumed (semi-)Markovian, and the focus is on modeling, representation of expected accumulated cash flows, and computation of certain transition (sub-)probabilities using modifications of Kolmgorov's forward differential equations.

In this paper, we consider general life insurance payment processes in a canonical jump process framework. We study the representation of the expected accumulated cash flows that arise when modifying the payments by scaling factors depending on specific jump times. Contrary to previous investigations in the actuarial literature with this focus, we do not impose any restrictions on the intertemporal dependence structure of the jump process. By using supermartingales as Radon-Nikodym derivatives, we find representations of these expected accumulated scaled cash flows on classic form w.r.t. a new probability measure. The probability measure is characterized in terms of the original measure and the scaling factors. In conjunction, these results shed light on the universality of the previously mentioned advances in the actuarial literature and provide a natural stepping stone for the derivation of efficient computation schemes beyond (semi-)Markovian models.

The paper is structured as follows. In Section 4.2, we motivate the investigation and present the probabilistic setup. Section 4.3 contains the main results. Proofs are given in Appendix 4.A.

4.2 Motivation and setup

Our work is motivated by recent advances in multi-state life insurance; this aspect is discussed in Subsection 4.2.1. In Subsections 4.2.2–4.2.3, we introduce and describe the general framework.

4.2.1 Motivation

In multi-state life insurance mathematics, key objects of interest include expected accumulated cash flows. If $B = (B(t))_{t\geq 0}$ is a suitably regular payment process and the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ constitutes the available information, then the corresponding expected accumulated cash flows A are given by

$$A(t,s) = \mathbb{E}[\,B(s) - B(t) \,|\, \mathcal{F}_t].$$

Markov chain models remain popular in both theory and practice, cf. Hoem (1969), Norberg (1991), and Buchardt and Møller (2015). In these models, the payments take the form

$$B(\mathrm{d}t) = \sum_{j\in\mathcal{J}} \mathbb{1}_{\{Z_t=j\}} b_j(t)\,\mathrm{d}t + \sum_{j,k\in\mathcal{J}:k\neq j} b_{jk}(t) N_{jk}(\mathrm{d}t) \tag{4.2.1}$$

for suitably regular deterministic sojourn payment rates b_j and transition payments b_{jk}. Here $Z = (Z_t)_{t\geq 0}$ is a Markovian jump process on a finite state space $\mathcal{J}$ admitting suitably regular transition rates μ. The compensators of the corresponding multivariate counting process $N = (N(t))_{t\geq 0}$ are then given by

$$\Lambda_{jk}(\mathrm{d}t) = \mathbb{1}_{\{Z_{t-}=j\}}\mu_{jk}(t)\,\mathrm{d}t.$$

The filtration $\mathcal{F}$ consists of the information naturally generated by Z. Since

$$A(t,\mathrm{d}s) \overset{\text{a.s.}}{=} \sum_{j\in\mathcal{J}} p_{Z_t j}(t,s)\left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j}\mu_{jk}(s)b_{jk}(s)\right)\mathrm{d}s$$

with p denoting the transition probabilities of Z, computation of the expected accumulated cash flow $A(t,\cdot)$ simply involves computation of the transition probabilities $p(t,\cdot)$ via Kolmogorov's forward differential equations.

In the last decade, the inclusion of incidental policyholder behavior has received significant interest, see e.g. Henriksen et al. (2014), Buchardt, Møller, and Schmidt (2015), Buchardt and Møller (2015), and Gad and Nielsen (2016). The inclusion of the free policy option and the option to retire earlier or later leads to payments that are scaled by a factor depending on the exercise time(s) of the option(s). This entails that the aforementioned forward method for the computation of expected accumulated cash flows appears to not be applicable.

To include incidental policyholder behavior, one may set $\mathcal{J} = \mathcal{J}_0 \cup \mathcal{J}_1$ and assume that $Z_t \in \mathcal{J}_1$ implies $Z_s \in \mathcal{J}_1$ for all $s > t$. With τ the first hitting time of $\mathcal{J}_1$, we can then interpret $\mathcal{J}_0$ as the states prior to exercise, τ as the exercise time, and $\mathcal{J}_1$ as the subsequent states. Interest now lies in payments B^ρ of the form

$$B^\rho(\mathrm{d}t) = \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau\leq t\}}}B(\mathrm{d}t)$$

with $\rho \in (0,1]$ some suitably regular scaling factor and B given by (4.2.1). By inspecting closely the methods and results of e.g. Buchardt and Møller (2015), it is possible to show that the expected accumulated cash flows are given by

$$A^\rho(t,\mathrm{d}s) \overset{\text{a.s.}}{=} \rho(\tau, Z_{\tau-})^{\mathbb{1}_{\{\tau\leq t\}}} \sum_{j\in\mathcal{J}} p^\rho_{Z_t j}(t,s)\left(b_j(s) + \sum_{k\in\mathcal{J}:k\neq j}\mu^\rho_{jk}(s)b_{jk}(s)\right)\mathrm{d}s,$$

where p^ρ are the transition probabilities of another Markovian jump process $Z^\rho = (Z^\rho_t)_{t\geq 0}$ with values in $\nabla \cup \mathcal{J}$ admitting transition rates μ^ρ given by

$$\begin{aligned}
\mu^\rho_{jk}(t) &= \rho(t,j)\mu_{jk}(t), && j\in\mathcal{J}_0, k\in\mathcal{J}_1,\\
\mu^\rho_{j\nabla}(t) &= (1-\rho(t,j))\sum_{k\in\mathcal{J}_1}\mu_{jk}(t), && j\in\mathcal{J}_0,\\
\mu^\rho_{j\nabla}(t) &= 0, && j\in\mathcal{J}_1,\\
\mu^\rho_{\nabla k}(t) &= 0, && k\in\mathcal{J},\\
\mu^\rho_{jk}(t) &= \mu_{jk}(t), && \text{otherwise.}
\end{aligned}$$

Computation of the expected accumulated cash flow $A^\rho(t,\cdot)$ thus involves computation of the transition probabilities p^ρ via Kolmogorov's forward differential equations. Consequently, the forward method for Markov chain models is easily adapted to take into account incidental policyholder behavior.

Besides the approach found in Christiansen and Djehiche (2020), which concerns backward methods and the determination of the scaling factors $\rho_1,\dots,\rho_n$ while maintaining actuarial equivalence, the literature focuses on (semi-)Markovian jump processes and at most two policyholder options. The above demonstration, which is akin to a change of measure, is new and actually alludes to a more general link between scaling factor and changes of measure; our focus is exactly on establishing this link. To this end, we investigate the general case consisting of an arbitrary (finite) number of policyholder options and no restrictions regarding the intertemporal dependence structure of the jump process. The provision of solutions to specific actuarial problems, in particular in the context of efficient computation of expected accumulated cash flows in the presence of free policy behavior and stochastic retirement, is postponed to future research. Correspondingly, the following presentation is shaped in a general probabilistic fashion, and the methods and results are aimed at users of multi-state models in general.

4.2.2 Canonical framework

Introduce the mark space $\mathcal{J} = \{\nabla\} \cup \mathcal{J}_0 \cup \cdots \cup \mathcal{J}_n$, $n \in \mathbb{N}$, with each $\mathcal{J}_i$ countable, equipped with the power-set $2^{\mathcal{J}}$. For $i = 1,\dots,n$ we denote by $\mathcal{J}_{i-}$ the set $\mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{i-1}$. For $i = 0,\dots,n-1$ we denote by $\mathcal{J}_{i+}$ the set $\mathcal{J}_{i+1} \cup \cdots \cup \mathcal{J}_n$. Consider a background probability space $(\Omega,\mathbb{F},\mathbb{P})$. Here $(\Omega,\mathbb{F})$ is taken to be the canonical measurable space of non-explosive random counting measures associated with the mark space $(\mathcal{J},2^{\mathcal{J}})$. We denote by ν° the canonical random counting measure given by the identity map from $(\Omega,\mathbb{F})$ onto $(\Omega,\mathbb{F})$. See Jacobsen (2006) for details regarding the canonical framework.

Let $(T,Y) = (T_n,Y_n)_{n\in\mathbb{N}}$ be a non-explosive marked point process with mark space $(\mathcal{J},2^{\mathcal{J}})$. We note that (T,Y) is an isomorphism (bijective and bimeasurable map) from $(\Omega,\mathbb{F})$ onto the canonical measurable space of non-explosive marked point processes with mark space $\mathcal{J}$.

We equip $(\Omega,\mathbb{F})$ with the canonical filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ generated by ν°. Recall that $\mathcal{F}$ is right-continuous. We are not going to complete $\mathcal{F}$; in other words, the *usual conditions* do not hold. Consequently, we rely on results pertaining to the canonical framework (as presented by e.g. Jacobsen, 2006) rather than results from 'the general theory of processes'.

Set $T_0 \equiv 0$ and $Y_0 \equiv z_0 \neq \nabla$, and let $Z = (Z_t)_{t\geq 0}$ be the non-explosive jump process associated with $(T_n, Y_n)_{n\in\mathbb{N}}$ In other words, Z is defined by

$$Z_t = Y_n \qquad (T_n \leq t < T_{n+1}).$$

Note that Z is not an isomorphism from $(\Omega, \mathbb{F})$ onto the canonical measurable space of non-explosive jump processes with values in $(\mathcal{J}, 2^{\mathcal{J}})$ since Z is not injective (identical successive marks are not identifiable from Z). Still, we might associate with Z a random counting measure $\nu : (\Omega, \mathbb{F}) \to (\Omega, \mathbb{F})$ such that $\nu([0,t] \times J)$ is the number of jumps of Z to $J \subset \mathcal{J}$ before time $t \geq 0$. Since Z is not injective, $\nu \neq \nu^\circ$.

It is natural and custom to specify the model through the (compensators of the) multivariate counting process $N = (N(t))_{t\geq 0}$ associated with Z given by $N_{jk}(0) = 0$ and

$$N_{jk}(t) = \#\{s \in (0,t] : Z_{s-} = j, Z_s = k\} = \int_{(0,t]} \mathbb{1}_{\{Z_{s-}=j\}}\, \nu(\mathrm{d}s \times \{k\})$$

for $j, k \in \mathcal{J}, k \neq j$, and $t > 0$. Observe that the natural filtrations $\mathcal{F}^Z$, $\mathcal{F}^N$, and $\mathcal{F}^\nu$ generated by Z, N, and ν, respectively, agree (see e.g. Jacobsen, 2006, p. 43 mid).

Our techniques rely on the canonical framework and thus pertain to the filtration $\mathcal{F}$. The probabilistic model can be specified via the canonical compensating measure C° w.r.t. $\mathbb{P}$, i.e. the compensating measure of ν° w.r.t. $(\mathcal{F}, \mathbb{P})$, cf. Jacobsen (2006, Section 4.8, in particular Theorem 4.8.1). Alternatively, one might want to specify the probabilistic model via the predictable compensators of N w.r.t. the natural filtration $\mathcal{F}^N$. In that case, the following observation concerning the relation between the canonical framework and the multivariate counting process N is important. When the model is specified via the multivariate counting process N, we can without loss of generality assume that there are no identical successive marks under $\mathbb{P}$, i.e.

$$\mathbb{P}(T_n < \infty, Y_n \neq Y_{n+1}) = \mathbb{P}(T_n < \infty) \tag{4.2.2}$$

for all $n \in \mathbb{N}_0$. This follows from the observations that the distribution of ν defines another probability measure on $(\Omega, \mathbb{F})$ that does not alter the distribution of N and for which (4.2.2) holds. When (4.2.2) holds, one can show that $\mathcal{F}$ and $\mathcal{F}^N$ only differ on $\mathbb{P}$-null sets and that ν and ν° are $\mathbb{P}$-indistinguishable. This ensures that we may specify the probabilistic model via the predictable compensators of N, alternatively via the compensating measure of ν, w.r.t. the natural filtration $\mathcal{F}^N$, rather than via the canonical compensating measure. In the following, we suppose (4.2.2) holds.

Terminology A real-valued stochastic process $X = (X_t)_{t\geq 0}$ is said to be continuous, càdlàg, of finite variation, etc., if that property holds for every path $X(\omega)$, $\omega \in \Omega$. Furthermore, X is said to be bounded if there exists a universal constant $K > 0$ such that $|X_t(\omega)| < K$ for all $\omega \in \Omega$ and $t \geq 0$. Also, a stochastic process X is said to be a finite variation process if X is real-valued, adapted to $\mathcal{F}$, of finite variation, and càdlàg. Finally, a finite variation process X is said to have $\mathbb{P}$-integrable variation if the variation $|X| = (|X|_t)_{t\geq 0}$ of X satisfies $\mathbb{E}[|X|_t] < \infty$ for all $t \geq 0$.

4.2.3 Stopping times and scaling factors

Define $\tau_1, \ldots, \tau_n$, and ζ as the first hitting times of $\mathcal{J}_1, \ldots, \mathcal{J}_n$, and $\{\nabla\}$ by Z, respectively:

$$\begin{aligned} \tau_i &:= \inf\{s \in [0, \infty) : Z_s \in \mathcal{J}_i\}, \qquad i = 1, \ldots, n, \\ \zeta &:= \inf\{s \in [0, \infty) : Z_s = \nabla\}. \end{aligned}$$

We use the convention $\inf \emptyset = \infty$. Note that τ and ζ are stopping times w.r.t. $\mathcal{F}$, see e.g. Jacobsen (2006, Proposition 4.2.1(b)(i)). The following assumptions are imposed:

Assumption 1: $\mathbb{P}(\zeta < \infty) = 0$. In other words, Z does not hit $\{\nabla\}$ under $\mathbb{P}$.

Assumption 2: For any $i = 1, \ldots, n$ and $j \in \mathcal{J}_i$, it holds that $\mathbb{P}(N_{jk}(t) > 0) = 0$ for all $k \in \mathcal{J}_{i-}$ and all $t \geq 0$.

In combination, Assumptions 1–2 in particular imply a decrement structure under $\mathbb{P}$ in the following sense: The jump process Z only exits $\mathcal{J}_i$ by a transition to $\mathcal{J}_{i+}$ whereafter return to $\mathcal{J}_{i-} \cup \mathcal{J}_i$ is impossible.

Let $\rho_1, \ldots, \rho_n$ be some real-valued $\mathcal{F}^N$-predictable and hence also $\mathcal{F}$-predictable processes. Assume $0 < \rho_i \leq 1$ and that ρ_i is bounded away from zero for each $i = 1, \ldots, n$.

Define the real-valued processes $H_i = (H_i(t))_{t\geq 0}$ by

$$H_i(t) = \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}}, \qquad t \geq 0,$$

for $i = 1, \ldots, n$, and define the real-valued process $H = (H(t))_{t\geq 0}$ by

$$H(t) = \prod_{i=1}^{n} H_i(t), \qquad t \geq 0.$$

It follows that $0 < H \leq 1$ bounded away from zero and that H is a finite variation process.

4.3 Main results

This section contains the main results of the paper. We show the existence of and characterize a probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathbb{F})$ such that for any finite variation process B with $\mathbb{P}$-integrable variation,

$$\mathbb{E}\left[\int_{(s,t]} H(u)\, B(\mathrm{d}u) \,\middle|\, \mathcal{F}_s\right] \overset{\mathbb{P}-\text{a.s.}}{=} \tilde{\mathbb{E}}\left[\int_{(s,t]} \mathbb{1}_{\{\zeta>u\}}\, B(\mathrm{d}u) \,\middle|\, \mathcal{F}_s\right] H(s) \tag{4.3.1}$$

for $0 \leq s < t < \infty$. Here and in the following, the operator $\tilde{\mathbb{E}}$ denotes $\tilde{\mathbb{P}}$-integration.

In Subsection 4.3.1, we construct the desired probability measure via the (conditional) finite-dimensional distributions of the corresponding marked point process. Formula (4.3.1) is proven in Subsection 4.3.2. Subsection 4.3.3 is concerned with the characterization of the new probability measure via its corresponding compensating measure.

4.3.1 Preliminaries and construction

In this subsection, we explicitly construct the desired probability measure $\tilde{\mathbb{P}}$ via the (conditional) finite-dimensional distributions of the corresponding marked point process. Before turning to the construction, we discuss a preliminary result concerning H. This discussion is intended to motivate the subsequent construction.

Suppose there exists a probability measure $\tilde{\mathbb{P}}$ such that (4.3.1) holds. Denote for $t \geq 0$ by $\mathbb{P}_t$ and $\tilde{\mathbb{P}}_t$ the restrictions of $\mathbb{P}$ and $\tilde{\mathbb{P}}$ to $\mathcal{F}_t$, respectively. We should then find that $\mathbb{P}_t \ll \tilde{\mathbb{P}}_t$, $t \geq 0$, with Radon-Nikodym derivatives (likelihood process) $\mathcal{L} = (\mathcal{L}_t)_{t\geq 0}$ given by

$$\mathcal{L}_t = \frac{H(0)}{H(t)} \mathbb{1}_{\{\zeta>t\}}. \tag{4.3.2}$$

We may ask: why expect the existence of a probability measure $\tilde{\mathbb{P}}$ yielding (4.3.2)? The following result helps to reveal what is going on behind the scenes:

Proposition 4.3.1. *H is a supermartingale w.r.t. $(\mathcal{F}, \mathbb{P})$.*

Proof. See Appendix 4.A. □

Remark 4.3.2. Unless $\rho_i = 1$ or $\mathbb{P}(\tau_i < \infty) = 0$ for all $i = 1, \ldots, n$, one actually obtains that H is a true supermartingale w.r.t. $(\mathcal{F}, \mathbb{P})$, confer e.g. with the proof of Proposition 4.3.1. ▽

According to Proposition 4.3.1, the process $\tilde{H} = (\tilde{H}(t))_{t\geq 0}$ defined by $\tilde{H}(t) = H(t)/H(0)$ is a supermartingale w.r.t. $(\mathcal{F}, \mathbb{P})$ satisfying $\mathbb{E}[\tilde{H}(0)] = 1$. Since $\tilde{H}$ is supermartingale, successful change of measure – in the sense of the likelihood process $\mathcal{L}$ being a (local) $\tilde{\mathbb{P}}$-martingale – requires 'losing' probability mass. Recall that due to Assumption 1, the jump process Z does not hit $\{\nabla\}$ under $\mathbb{P}$, and consequently, such a loss might be achieved by giving positive probability to hitting $\{\nabla\}$. This idea is also reflected in (4.3.2). The resulting probability measure $\tilde{\mathbb{P}}$ would be a so-called

Föllmer measure associated with the supermartingale $\tilde{H}$, and $(\tilde{\mathbb{P}}, \zeta)$ would be a so-called *Föllmer pair* for the supermartingale $\tilde{H}$ (see Definition 2.1 in Perkowski and Ruf, 2015). General existence and non-uniqueness results for Föllmer measures are given in Perkowski and Ruf (2015).

In the following, we explicitly construct a Föllmer measure for $\tilde{H}$ starting from the (conditional) finite-dimensional distributions of the corresponding marked point process. The specific choice of new (conditional) finite-dimensional distributions might seem rather unmotivated, but it is actually inspired by (4.3.2) using the close link between likelihood processes, compensating measures for the random counting measures, and (conditional) finite-dimensional distributions for the corresponding marked point processes (see e.g. Jacod, 1975; Jacobsen, 2006). In other words, it results from an act of reverse engineering starting from (4.3.2).

Denote with $\langle t- \rangle$ the total number of jumps up until but not including time $t \in (0, \infty)$, i.e.

$$\langle t- \rangle := \nu^\circ([0, t) \times \mathcal{J}),$$

and denote with Ξ_m, $m \in \mathbb{N}_0$, the first m jump times and marks, i.e.

$$\Xi_m := (T_1, Y_1, \dots, T_m, Y_m)$$

with $(T_0, Y_0) \equiv (0, z_0)$. Further, denote with P^m, $m \in \mathbb{N}_0$, the regular conditional distribution of

$$T_{m+1} \quad \text{given} \quad \Xi_m \quad \text{under} \quad \mathbb{P},$$

and denote with π^m, $m \in \mathbb{N}_0$, the regular conditional distribution of

$$Y_{m+1} \quad \text{given} \quad (\Xi_m, T_{m+1}) \quad \text{under} \quad \mathbb{P}.$$

In particular, P^0 is the distribution of T_1 under $\mathbb{P}$ and π^0 is the conditional distribution of Y_1 given T_1 under $\mathbb{P}$. In the remainder of the paper, we consider a fixed version of these regular conditional distributions. Note that we hereby also implicitly fix a version of the corresponding compensating measure and a version of the corresponding predictable compensators (cf. Jacobsen, 2006, Subsection 4.3).

We now turn our attention to the construction of the new probability measure. Since the processes $\rho_1, \dots, \rho_n$ are assumed predictable, there exists measurable functions $(\Xi_m, t) \mapsto f_i^m(\Xi_m, t) \in (0, 1]$, $m \in \mathbb{N}_0$, such that

$$\rho_i(t) = f_i^{\langle t- \rangle}(\Xi_{\langle t- \rangle}, t)$$

for $t \geq 0$ using the convention $\langle 0- \rangle = 0$. For details we refer to e.g. Jacobsen (2006, Section 4.2, in particular Proposition 4.2.1(b)(iv)). We may then define new regular conditional distributions $\tilde{P}^m$ of T_{m+1} given Ξ_m and new regular conditional distributions of $\tilde{\pi}^m$ of Y_{m+1} given (Ξ_m, T_{m+1}) by setting $\tilde{P}^0 = P^0$ and

$$\tilde{P}^m([0,t] \,|\, \Xi_m) = \begin{cases} 0, & \nabla \in \{Y_1, \ldots, Y_m\}, \\ P^m([0,t] \,|\, \Xi_m), & \text{otherwise}, \end{cases}$$

for $m \in \mathbb{N}$, as well as setting

$$\tilde{\pi}^m(\{k\} \,|\, \Xi_m, T_{m+1}) = f_i^m(\Xi_m, T_{m+1})^{\mathbb{1}_{\{Y_m \in \mathcal{J}_{i-}\}}} \pi^m(\{k\} \,|\, \Xi_m, T_{m+1}),$$

for $k \in \mathcal{J}_i$, $i = 0, \ldots, n$, and $m \in \mathbb{N}_0$ (using the convention $\mathbb{1}_{\{Y_m \in \mathcal{J}_{0-}\}} = 0$) and

$$\tilde{\pi}^m(\{\nabla\} \,|\, \Xi_m, T_{m+1}) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_m \in \mathcal{J}_{i-}\}}(1 - f_i^m(\Xi_m, T_{m+1})) \sum_{k \in \mathcal{J}_i} \pi^m(\{k\} \,|\, \Xi_m, T_{m+1}),$$

for $m \in \mathbb{N}_0$. Note that for $k \neq \nabla$,

$$\tilde{\pi}^m(\{k\} \,|\, \Xi_m, T_{m+1}) \leq \pi^m(\{k\} \,|\, \Xi_m, T_{m+1}) \tag{4.3.3}$$

since $\rho_i \leq 1$ for $i = 1, \ldots, n$.

In the following, we consider the above version of the regular conditional distributions fixed. Note that we hereby also implicitly fix a version of the corresponding compensating measure and a version of the corresponding predictable compensators. Setting $\tilde{P}^m([0,t] \,|\, \Xi_m) = 0$ on $(Y_m = \nabla)$ is not necessary; since Z does not hit $\{\nabla\}$ under $\mathbb{P}$, the behavior of Z after hitting $\{\nabla\}$ under the new probability measure is not important for the result we develop. But it does ensure some type of minimality of the version we fix, in the sense that $\{\nabla\}$ becomes absorbing under the new probability measure. The antecedent discussion is related to non-uniqueness of Föllmer measures, see also Perkowski and Ruf (2015).

An application of the Ionescu-Tulcea theorem now yields a uniquely defined probability measure $\bar{\mathbb{P}}$ on the canonical measurable space of possibly explosive marked point processes with mark space $\mathcal{J}$ with (conditional) marginals $\tilde{P}^m$ and $\tilde{\pi}^m$. It essentially only remains to be shown that we can restrict $\bar{\mathbb{P}}$ to the canonical measurable space of non-explosive marked point processes.

To verify that the restriction is possible, we need to establish the identity

$$\bar{\mathbb{P}}\Big[\lim_{m\to\infty} T_m = \infty\Big] = 1.$$

Since $(T_m)_{m\in\mathbb{N}}$ is increasing, this is equivalent to

$$\forall t > 0 : \lim_{m\to\infty} \bar{\mathbb{P}}[T_m \leq t] = 0.$$

Recall that $(T_m)_{m\in\mathbb{N}}$ is non-explosive under $\mathbb{P}$. It then holds for all $t > 0$ that

$$\lim_{m\to\infty} \mathbb{P}[T_m \leq t] = 0,$$

and we can conclude that it is sufficient to establish the identity

$$\bar{\mathbb{P}}[T_m \leq t] \leq \mathbb{P}[T_m \leq t]$$

for all $t > 0$ and $m \in \mathbb{N}$.

Lemma 4.3.3. *For all $t > 0$ and $m \in \mathbb{N}$,*

$$\bar{\mathbb{P}}[T_m \leq t] \leq \mathbb{P}[T_m \leq t].$$

Proof. See Appendix 4.A. □

Collecting results, we conclude that it is possible to restrict $\bar{\mathbb{P}}$ to the canonical measurable space of non-explosive marked point processes. Consequently the inverse of the isomorphism (T, Y) induces a new probability measure $\tilde{\mathbb{P}}$ on $(\Omega, \mathbb{F})$ under which (T, Y) has (conditional) marginals $\tilde{P}^m$ and $\tilde{\pi}^m$ for $m \in \mathbb{N}_0$.

4.3.2 Expectation formulas

We now turn our attention to establishing (4.3.1). Recall that the new probability measure $\tilde{\mathbb{P}}$ was constructed exactly with (4.3.2) in mind. The following result confirms the intention of the construction.

Proposition 4.3.4. *For any $t \geq 0$ it holds that $\mathbb{P}_t \ll \tilde{\mathbb{P}}_t$ with Radon-Nikodym derivative*

$$\mathcal{L}_t := \frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\tilde{\mathbb{P}}_t} = \frac{1}{\tilde{H}(t)} \mathbb{1}_{\{\zeta > t\}},$$

which defines a bounded càdlàg martingale $\mathcal{L} = (\mathcal{L}_t)_{t \geq 0}$ w.r.t. $(\mathcal{F}, \tilde{\mathbb{P}})$.

Proof. Local absolute continuity and the fact that $\mathcal{L}$ is a càdlàg martingale w.r.t. $(\mathcal{F}, \tilde{\mathbb{P}})$ follows immediately by an application of Jacobsen (2006, Theorem 5.1.1(b)). Since the assumptions on $\rho_1, \ldots, \rho_n$ guarantee that $\tilde{H}$ is bounded away from zero, $\mathcal{L}$ is also bounded. □

By invoking the Radon-Nikodym derivatives (likelihood process) of Proposition 4.3.4, we immediately arrive at (4.3.1).

Theorem 4.3.5. *Let $B = (B(t))_{t \geq 0}$ be a finite variation process with $\mathbb{P}$-integrable variation. Then for any $t > 0$,*

$$\mathbb{E}\left[\int_{(s,t]} H(u)\, B(\mathrm{d}u) \,\middle|\, \mathcal{F}_s \right] \overset{\mathbb{P}-a.s.}{=} \tilde{\mathbb{E}}\left[\int_{(s,t]} \mathbb{1}_{\{\zeta > u\}}\, B(\mathrm{d}u) \,\middle|\, \mathcal{F}_s \right] H(s), \qquad 0 \leq s < t.$$

In particular,

$$\mathbb{E}\left[\int_{(0,t]} H(u)\, B(\mathrm{d}u) \right] = \tilde{\mathbb{E}}\left[\int_{(0,t]} \mathbb{1}_{\{\zeta > u\}}\, B(\mathrm{d}u) \right] H(0).$$

Proof. See Appendix 4.A. □

With B defined by $B(s) = \mathbb{1}_{\{s \geq t\}} X$, $s \geq 0$, for a given $t \geq 0$ and some real-valued, $\mathcal{F}_t$-measurable, and $\mathbb{P}$-integrable random variable X, an application of Theorem 4.3.5 yields the following corollary:

Corollary 4.3.6. *Let $t \geq 0$, and let X be a real-valued, $\mathcal{F}_t$-measurable, and $\mathbb{P}$-integrable random variable. Then*

$$\mathbb{E}\big[\, H(t)X \,\big|\, \mathcal{F}_s \big] \overset{\mathbb{P}-a.s.}{=} \tilde{\mathbb{E}}\big[\mathbb{1}_{\{\zeta > t\}} X \,\big|\, \mathcal{F}_s \big]\, H(s), \qquad 0 \leq s \leq t.$$

In particular,

$$\mathbb{E}\Big[\tilde{H}(t)X\Big] = \tilde{\mathbb{E}}\big[\mathbb{1}_{\{\zeta > t\}} X\big].$$

Remark 4.3.7. Corollary 4.3.6 confirms that $(\tilde{\mathbb{P}}, \zeta)$ is indeed a Föllmer pair for the $\mathbb{P}$-supermartingale $\tilde{H}$, cf. Perkowski and Ruf (2015, Proposition 2.3). ▽

Example 4.3.8 (True martingales)**.** The following condition, which e.g. ensures that the differences $N - \Lambda$ are true (rather than only local) martingales w.r.t. $(\mathcal{F}, \mathbb{P})$, is often imposed:

$$\mathbb{E}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} N_{jk}(t)\Bigg] < \infty, \qquad t \geq 0. \tag{4.3.4}$$

Suppose (4.3.4) holds. Since $\{\nabla\}$ is absorbing under $\tilde{\mathbb{P}}$, it follows from Theorem 4.3.5 and the inequality $H \leq 1$ that

$$\begin{aligned}
\tilde{\mathbb{E}}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} N_{jk}(t)\Bigg] H(0) &= \tilde{\mathbb{E}}\Bigg[\sum_{\substack{j \in \mathcal{J} \\ j \neq \nabla}} N_{j\nabla}(t)\Bigg] H(0) + \tilde{\mathbb{E}}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} \int_{(0,t]} \mathbb{1}_{\{\zeta > u\}}\, N_{jk}(\mathrm{d}u)\Bigg] H(0) \\
&\leq H(0) + \mathbb{E}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} \int_{(0,t]} H(u)\, N_{jk}(\mathrm{d}u)\Bigg] \\
&\leq 1 + \mathbb{E}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} N_{jk}(t)\Bigg]
\end{aligned}$$

for any $t > 0$. Since H is positive and bounded away from zero, we conclude that (4.3.4) implies

$$\tilde{\mathbb{E}}\Bigg[\sum_{\substack{j,k \in \mathcal{J} \\ k \neq j}} N_{jk}(t)\Bigg] < \infty, \qquad t \geq 0.$$

In other words, conditions such as (4.3.4) are stable w.r.t. change of measure from $\mathbb{P}$ to $\tilde{\mathbb{P}}$. ○

4.3.3 Characterization

We conclude our study by characterizing the new probability measure $\tilde{\mathbb{P}}$ in terms of its corresponding compensating measure.

The canonical compensating measure C° w.r.t. $\mathbb{P}$, i.e. the compensating measure of ν° w.r.t. $(\mathcal{F}, \mathbb{P})$, is given by (for $J \in 2^{\mathcal{J}}$)

$$C^\circ(\mathrm{d}t \times J) = \pi^{\langle t- \rangle}(J \,|\, \Xi_{\langle t- \rangle}, t) \frac{P^{\langle t- \rangle}(\mathrm{d}t \,|\, \Xi_{\langle t- \rangle})}{1 - P^{\langle t- \rangle}(t- \,|\, \Xi_{\langle t- \rangle})},$$

cf. Jacobsen (2006, Definition 4.3.2), and the predictable compensators Λ of N w.r.t. $(\mathcal{F}, \mathbb{P})$ are given by

$$\Lambda_{jk}(\mathrm{d}t) = \mathbb{1}_{\{Z_{t-} = j\}} \, C^\circ(\mathrm{d}t \times \{k\}).$$

Denote by $\tilde{C}^\circ$ the canonical compensating measure w.r.t. $\tilde{\mathbb{P}}$, i.e. the compensating measure of ν° w.r.t. $(\mathcal{F}, \tilde{\mathbb{P}})$, and denote by $\tilde{\Lambda}$ the predictable compensators of N w.r.t. $(\mathcal{F}, \tilde{\mathbb{P}})$.

Proposition 4.3.9. *It holds that*

$$\tilde{C}^\circ(\mathrm{d}t \times J) = \mathbb{1}_{\{\varsigma \geq t\}} \rho_i(t)^{\mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}}} C^\circ(\mathrm{d}t \times J), \quad J \subset \mathcal{J}_i, i = 0, \ldots, n,$$

$$\tilde{C}^\circ(\mathrm{d}t \times \{\nabla\}) = \mathbb{1}_{\{\varsigma \geq t\}} \sum_{i=1}^{n} \mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}} (1 - \rho_i(t)) \, C^\circ(\mathrm{d}t \times \mathcal{J}_i),$$

where we use the convention $\mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{0-}\}} = 0$. *In particular,*

$$\begin{aligned}
\tilde{\Lambda}_{jk}(\mathrm{d}t) &= \mathbb{1}_{\{\varsigma \geq t\}} \rho_\ell(t) \, \Lambda_{jk}(\mathrm{d}t), && j \in \mathcal{J}_i, k \in \mathcal{J}_\ell, \ell > i, \\
& && i = 0, \ldots, n-1, \\
\tilde{\Lambda}_{j\nabla}(\mathrm{d}t) &= \mathbb{1}_{\{\varsigma \geq t\}} \sum_{\ell = i+1}^{n} (1 - \rho_\ell(t)) \sum_{k \in \mathcal{J}_\ell} \Lambda_{jk}(\mathrm{d}t), && j \in \mathcal{J}_i, i = 0, \ldots, n-1, \\
\tilde{\Lambda}_{\nabla k}(\mathrm{d}t) &= 0, && k \in \mathcal{J}, k \neq \nabla, \\
\tilde{\Lambda}_{jk}(\mathrm{d}t) &= \mathbb{1}_{\{\varsigma \geq t\}} \Lambda_{jk}(\mathrm{d}t), && \textit{otherwise.}
\end{aligned}$$

Proof. See Appendix 4.A. □

Example 4.3.10 (Markovian jump processes)**.** Suppose that Z is a Markovian jump process admitting suitably regular transition rates μ under $\mathbb{P}$, and suppose that $\rho_i(t) = g_i(t, Z_{t-})$ for suitably regular deterministic functions $g_1, \ldots, g_n$. Since the predictable compensators characterize the distribution of the jump process Z, Proposition 4.3.9 yields that Z is also Markovian under $\tilde{\mathbb{P}}$, but instead admits transition rates $\tilde{\mu}$ of the form

$$\begin{aligned}
\tilde{\mu}_{jk}(t) &= g_\ell(t,j)\mu_{jk}(t), && j \in \mathcal{J}_i, k \in \mathcal{J}_\ell, \ell > i, i = 0,\dots,n-1,\\
\tilde{\mu}_{j\nabla}(t) &= \sum_{\ell=i+1}^{n} (1 - g_\ell(t,j)) \sum_{k \in \mathcal{J}_\ell} \mu_{jk}(t), && j \in \mathcal{J}_i, i = 0,\dots,n-1,\\
\tilde{\mu}_{\nabla k}(t) &= 0, && k \in \mathcal{J}, k \neq \nabla,\\
\tilde{\mu}_{jk}(t) &= \mu_{jk}(t), && \text{otherwise.}
\end{aligned}$$

This characterization in combination with Theorem 4.3.5 can be used to generalize the methods and results presented in Subsection 4.2.1 to an arbitrary (finite) number of policyholder options. ◦

Acknowledgments

Christian Furrer's research is partly funded by the Innovation Fund Denmark under File No. 7038-00007B. We should like to thank Martin Jacobsen for fruitful discussions and suggestions concerning Lemma 4.3.3.

4.A Proofs

Proof of Proposition 4.3.1. We give a proof by induction. We first consider the base case with $n = 1$. For all $0 \leq s \leq t < \infty$ we find almost surely w.r.t. $\mathbb{P}$ that

$$\begin{aligned}
\mathbb{E}\big[\rho_1(\tau_1)^{\mathbb{1}_{\{\tau_1 \leq t\}}} \,\big|\, \mathcal{F}_s\big] &= \mathbb{1}_{\{\tau_1 \leq s\}}\rho_1(\tau_1) + \mathbb{1}_{\{\tau_1 > s\}}\mathbb{E}\big[\rho_1(\tau_1)^{\mathbb{1}_{\{\tau_1 \leq t\}}} \,\big|\, \mathcal{F}_s\big]\\
&\leq \mathbb{1}_{\{\tau_1 \leq s\}}\rho_1(\tau_1) + \mathbb{1}_{\{\tau_1 > s\}}\\
&= \rho_1(\tau_1)^{\mathbb{1}_{\{\tau_1 \leq s\}}},
\end{aligned}$$

which shows that H is a supermartingale w.r.t. $(\mathcal{F}, \mathbb{P})$ in the base case. For the induction step, we assume the supermartingale property w.r.t. $(\mathcal{F}, \mathbb{P})$ holds for some $k \in \mathbb{N}$. First note that for all $0 \leq s \leq t < \infty$,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{k+1} \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}} \,\middle|\, \mathcal{F}_s\right] &= \mathbb{E}\left[\rho_{k+1}(\tau_{k+1})^{\mathbb{1}_{\{\tau_{k+1} \leq t\}}} \prod_{i=1}^{k} \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}} \,\middle|\, \mathcal{F}_s\right]\\
&= \mathbb{1}_{\{\tau_{k+1} \leq s\}}\rho_{k+1}(\tau_{k+1})\mathbb{E}\left[\prod_{i=1}^{k} \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}} \,\middle|\, \mathcal{F}_s\right]\\
&\quad + \mathbb{1}_{\{\tau_{k+1} > s\}}\mathbb{E}\left[\rho_{k+1}(\tau_{k+1})^{\mathbb{1}_{\{\tau_{k+1} \leq t\}}} \prod_{i=1}^{k} \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}} \,\middle|\, \mathcal{F}_s\right]\\
&\leq \rho_{k+1}(\tau_{k+1})^{\mathbb{1}_{\{\tau_{k+1} \leq s\}}}\mathbb{E}\left[\prod_{i=1}^{k} \rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i \leq t\}}} \,\middle|\, \mathcal{F}_s\right]
\end{aligned}$$

almost surely w.r.t. $\mathbb{P}$. From the induction hypothesis we may then conclude that

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{k+1}\rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i\leq t\}}}\,\middle|\,\mathcal{F}_s\right] &\leq \rho_{k+1}(\tau_{k+1})^{\mathbb{1}_{\{\tau_{k+1}\leq s\}}}\prod_{i=1}^{k}\rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i\leq s\}}}\\ &= \prod_{i=1}^{k+1}\rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i\leq s\}}}\end{aligned}$$

almost surely. Consequently, the real-valued stochastic process $[0,\infty) \ni t \mapsto \prod_{i=1}^{k+1}\rho_i(\tau_i)^{\mathbb{1}_{\{\tau_i\leq t\}}}$ is a supermartingale w.r.t. $(\mathcal{F},\mathbb{P})$. This completes the induction step and thus the proof. $\square$

Proof of Lemma 4.3.3. Fix $t > 0$ and $m \in \mathbb{N}$. Since Z does not hit $\{\nabla\}$ under $\mathbb{P}$ and $\{\nabla\}$ is absorbing under $\bar{\mathbb{P}}$,

$$\begin{aligned}\bar{\mathbb{P}}[T_{m+1}\leq t] &= \bar{\mathbb{P}}[T_{m+1}\leq t, Y_1\neq\nabla,\ldots,Y_m\neq\nabla],\\ \mathbb{P}[T_{m+1}\leq t] &= \mathbb{P}[T_{m+1}\leq t, Y_1\neq\nabla,\ldots,Y_m\neq\nabla].\end{aligned}$$

Straightforward calculations with $\xi_i = (t_1, y_1, \ldots, t_i, y_i)$ for $i = 1,\ldots,m$ then yield

$$\begin{aligned}\bar{\mathbb{P}}[T_{m+1}\leq t] = \sum_{y_1,\ldots,y_m\neq\nabla}\int_0^t\cdots\int_0^t\Big\{&\tilde{P}^{m-1}(\mathrm{d}t_m\,|\,\xi_{m-1})\cdots\tilde{P}^0(\mathrm{d}t_1)\\ &\tilde{P}^m([0,t]\,|\,\xi_m)\prod_{i=1}^m\tilde{\pi}^{i-1}(\{y_i\}\,|\,\xi_{i-1},t_i)\Big\},\\ \mathbb{P}[T_{m+1}\leq t] = \sum_{y_1,\ldots,y_m\neq\nabla}\int_0^t\cdots\int_0^t\Big\{&P^{m-1}(\mathrm{d}t_m\,|\,\xi_{m-1})\cdots P^0(\mathrm{d}t_1)\\ &P^m([0,t]\,|\,\xi_m)\prod_{i=1}^m\pi^{i-1}(\{y_i\}\,|\,\xi_{i-1},t_i)\Big\}.\end{aligned}$$

Referring to the definition of $\tilde{P}^i$ and $\tilde{\pi}^i$ from P^i and π^i as well as the inequality (4.3.3), we may thus conclude that

$$\bar{\mathbb{P}}[T_{m+1}\leq t]\leq\mathbb{P}[T_{m+1}\leq t]$$

as desired. $\square$

Proof of Theorem 4.3.5. Fix $t > 0$ and $s \in [0,t)$. Note that $\int_{(s,t]} H(u)\,B(\mathrm{d}u)$ is $\mathcal{F}_t$-measurable and $\mathbb{P}$-integrable. Recall from Proposition 4.3.4 that $\mathbb{P}_t \ll \tilde{\mathbb{P}}_t$ with Radon-Nikodym derivative

$$\mathcal{L}_t := \frac{\mathrm{d}\mathbb{P}_t}{\mathrm{d}\tilde{\mathbb{P}}_t} = \frac{1}{\tilde{H}(t)}\mathbb{1}_{\{\zeta>t\}} = \frac{H(0)}{H(t)}\mathbb{1}_{\{\zeta>t\}}.$$

Then by Bayes' theorem,

$$\mathbb{E}\left[\int_{(s,t]} H(u)\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right] \overset{\mathbb{P}-\text{a.s.}}{=} \tilde{\mathbb{E}}\left[\mathcal{L}_t \int_{(s,t]} H(u)\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right]\frac{H(s)}{H(0)},$$

since $\mathbb{P}(\zeta < \infty) = 0$ by Assumption 1. Let $F \in \mathcal{F}_s$. Similar to Jacod and Shiryaev (2003, proof of Lemma I.3.12), we find that

$$\begin{aligned}\tilde{\mathbb{E}}\left[\mathbb{1}_F\,\mathcal{L}_t \int_{(s,t]} H(u)\,B(\mathrm{d}u)\right] &= \tilde{\mathbb{E}}\left[\mathbb{1}_F \int_{(s,t]} \mathcal{L}_u\,H(u)\,B(\mathrm{d}u)\right] \\ &= \tilde{\mathbb{E}}\left[\mathbb{1}_F \int_{(s,t]} \mathbb{1}_{\{\zeta>u\}}\,B(\mathrm{d}u)\right]H(0).\end{aligned}$$

It follows that

$$\tilde{\mathbb{E}}\left[\mathcal{L}_t \int_{(s,t]} H(u)\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right] \overset{\tilde{\mathbb{P}}-\text{a.s.}}{=} \tilde{\mathbb{E}}\left[\int_{(s,t]} \mathbb{1}_{\{\zeta>u\}}\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right]H(0),$$

where the equality also holds $\mathbb{P}$-a.s. due to the fact that $\mathbb{P}_s \ll \tilde{\mathbb{P}}_s$. All in all,

$$\mathbb{E}\left[\int_{(s,t]} H(u)\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right] \overset{\mathbb{P}-\text{a.s.}}{=} \tilde{\mathbb{E}}\left[\int_{(s,t]} \mathbb{1}_{\{\zeta>u\}}\,B(\mathrm{d}u)\,\middle|\,\mathcal{F}_s\right]H(s),$$

as desired. In particular, since $\mathcal{F}_0$ is trivial, we have

$$\mathbb{E}\left[\int_{(0,t]} H(u)\,B(\mathrm{d}u)\right] = \tilde{\mathbb{E}}\left[\int_{(0,t]} \mathbb{1}_{\{\zeta>u\}}\,B(\mathrm{d}u)\right]H(0),$$

which completes the proof. □

Proof of Proposition 4.3.9. The statements regarding the predictable compensators of N follow immediately from the statements regarding the canonical compensating measure. We thus focus on the latter statements. Referring to Jacobsen (2006, Definition 4.3.2),

$$\tilde{C}^{\circ}(\mathrm{d}t \times J) = \tilde{\pi}^{\langle t-\rangle}(J \,|\, \Xi_{\langle t-\rangle}, t)\frac{\tilde{P}^{\langle t-\rangle}(\mathrm{d}t \,|\, \Xi_{\langle t-\rangle})}{1 - \tilde{P}^{\langle t-\rangle}(t- \,|\, \Xi_{\langle t-\rangle})},$$

for $J \in 2^{\mathcal{J}}$. Note that according to the construction of $\tilde{\mathbb{P}}$,

$$\frac{\tilde{P}^{\langle t-\rangle}(\mathrm{d}t \,|\, \Xi_{\langle t-\rangle})}{1 - \tilde{P}^{\langle t-\rangle}(t- \,|\, \Xi_{\langle t-\rangle})} = \mathbb{1}_{\{\zeta \geq t\}}\frac{P^{\langle t-\rangle}(\mathrm{d}t \,|\, \Xi_{\langle t-\rangle})}{1 - P^{\langle t-\rangle}(t- \,|\, \Xi_{\langle t-\rangle})}.$$

If $J \subset \mathcal{J}_i$, $i = 0, \ldots, n$, then also according to the construction of $\tilde{\mathbb{P}}$,

$$\begin{aligned}\tilde{\pi}^{\langle t-\rangle}(J \,|\, \Xi_{\langle t-\rangle}, t) &= f_i^{\langle t-\rangle}(\Xi_{\langle t-\rangle}, t)^{\mathbb{1}_{\{Y_{\langle t-\rangle} \in \mathcal{J}_{i-}\}}}\pi^{\langle t-\rangle}(J \,|\, \Xi_{\langle t-\rangle}, t) \\ &= \rho_i(t)^{\mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}}}\pi^{\langle t-\rangle}(J \,|\, \Xi_{\langle t-\rangle}, t).\end{aligned}$$

In particular,

$$\begin{aligned}\tilde{C}^\circ(\mathrm{d}t \times J) &= \mathbb{1}_{\{\zeta \geq t\}} \rho_i(t)^{\mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}}} \pi^{\langle t- \rangle}(J \,|\, \Xi_{\langle t- \rangle}, t) \frac{P^{\langle t- \rangle}(\mathrm{d}t \,|\, \Xi_{\langle t- \rangle})}{1 - P^{\langle t- \rangle}(t- \,|\, \Xi_{\langle t- \rangle})} \\ &= \mathbb{1}_{\{\zeta \geq t\}} \rho_i(t)^{\mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}}} C^\circ(\mathrm{d}t \times J).\end{aligned}$$

In similar fashion,

$$\begin{aligned}&\tilde{\pi}^{\langle t- \rangle}(\{\nabla\} \,|\, \Xi_{\langle t- \rangle}, t) \\ &= \sum_{i=1}^n \mathbb{1}_{\{Y_{\langle t- \rangle} \in \mathcal{J}_{i-}\}} \Big(1 - f_i^{\langle t- \rangle}(\Xi_{\langle t- \rangle}, t)\Big) \sum_{k \in \mathcal{J}_i} \pi^{\langle t- \rangle}(\{k\} \,|\, \Xi_{\langle t- \rangle}, t) \\ &= \sum_{i=1}^n \mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}} (1 - \rho_i(t)) \sum_{k \in \mathcal{J}_i} \pi^{\langle t- \rangle}(\{k\} \,|\, \Xi_{\langle t- \rangle}, t).\end{aligned}$$

In particular,

$$\begin{aligned}&\tilde{C}^\circ(\mathrm{d}t \times \nabla) \\ &= \mathbb{1}_{\{\zeta \geq t\}} \sum_{i=1}^n \mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}} (1 - \rho_i(t)) \sum_{k \in \mathcal{J}_i} \pi^{\langle t- \rangle}(\{k\} \,|\, \Xi_{\langle t- \rangle}, t) \frac{P^{\langle t- \rangle}(\mathrm{d}t \,|\, \Xi_{\langle t- \rangle})}{1 - P^{\langle t- \rangle}(t- \,|\, \Xi_{\langle t- \rangle})} \\ &= \mathbb{1}_{\{\zeta \geq t\}} \sum_{i=1}^n \mathbb{1}_{\{Z_{t-} \in \mathcal{J}_{i-}\}} (1 - \rho_i(t))\, C^\circ(\mathrm{d}t \times \mathcal{J}_i).\end{aligned}$$

Collecting results completes the proof. $\square$

Chapter 5

Forward transition rates

This chapter contains the paper *Buchardt, Furrer, and Steffensen (2019).*

Abstract

The idea of forward rates stems from interest rate theory. It has natural connotations to transition rates in multi-state models. The generalization from the forward mortality rate in a survival model to multi-state models is non-trivial and several definitions have been proposed. We establish a theoretical framework for the discussion of forward rates. Furthermore, we provide a novel definition with its own logic and merits and compare it with the proposals in the literature. The definition turns the Kolmogorov forward equations inside out by interchanging the transition probabilities with the transition intensities as the object to be calculated.

Keywords: Forward rates; Doubly stochastic Markov models; Life insurance; Kolmogorov forward equations

5.1 Introduction

We provide a novel concept of forward transition rates in multi-state models with applications to life insurance as well as credit risk. It is a purely probabilistic concept that is tailor-made to match transition probabilities in a specific way, even in state models that are not Markovian. Though simple and constructive, our forward transition rates are different from the ones suggested in the literature, mainly with applications to life insurance in mind. Our contribution is three-fold. We propose a novel multi-state definition, we analyze its characteristics, and we compare these characteristics with those of other definitions proposed earlier.

Forward interest rates play an important role in bond market theory. They allow us to represent, at a fixed time point, both prices of nominal payments and prices of future interest rates 'as if' the future interest rates were known and equal to the forward rates. The forward interest rate curve has even been considered as the fundamental object to model, rather than the interest rate, by a stochastic infinite-dimensional process with certain consistency constraints.

Two areas of finance and insurance that are closely linked, at least from a probabilistic point of view, are (reduced form) credit risk theory and life insurance mathematics. The relation between the areas has been explored and exploited by e.g. Kraft and Steffensen (2007). In both disciplines, the doubly stochastic finite-state Markov chain is a fundamental stochastic model. The health and life status of an insured (or, in credit risk theory, the credit rating and solvency status of a firm) is modeled as a finite-state chain. In the doubly stochastic Markov setting, this finite-state chain is assumed to be Markov, conditional on the transition rates. These rates depend on macro-demographic conditions in the population (or, in credit risk theory, macro-economic conditions in the market and the political regime).

When studying transition probabilities and transition densities in these models, the relation to forward rates in bond market theory is striking – particularly in the simple survival model, where there are only two states of which one is absorbing. This was first observed and exploited by Milevsky and Promislow (2001), while Miltersen and Persson (2005) discussed the extension to stochastic interest rates, allowing for correlation between mortality and interest. The idea about generalization to multi-state models was discussed and researched by several academics during those years but were put on halt by Norberg (2010) who explained thoroughly the drawback of each and every natural generalizing definition. As it turned out, this was not enough to quell the idea in and of itself. Lately, Christiansen and Niemeyer (2015) and Buchardt (2017) have proposed different generalizations with individual characteristics.

We propose here yet another generalizing definition with its own logic and merits. It is based on the simple idea to consider the system of Kolmogorov forward equations not as a means of calculating transition probabilities for given transition rates, but instead as a means of calculating transition rates for given transition probabilities. One version of the idea was already considered by Norberg (2010) but rejected due to general non-uniqueness of the solution, i.e. non-uniqueness of the forward transition rates. But Norberg's version took the initial state for given. If the equations have to hold for a portfolio of insured (or a portfolio of firms in the credit risk version), distributed over the state space at the starting time point, we obtain far stronger results regarding existence and uniqueness.

The definition has drawbacks in specific applications to insurance, though. The transition probabilities arising from our forward transition rates in an 'assumed to be' Markovian setting are actually the correct transition probabilities – this is exactly how they are constructed. However, the transition probabilities and our forward transition rates do not form together, in an 'assumed to be' Markovian setting, the densities of transitions. This limits their application for calculation of relevant actuarial quantities. We indicate, however, how this drawback partly can be compensated by extending the model artificially.

One part of our contribution is our specific proposal. Another part of our contribution is the establishment of a theoretical framework for the discussion and comparison of forward rate definitions. This allows us to give a clear presentation of the relation between the different suggestions pushed forward by Christiansen and Niemeyer (2015), Buchardt (2017), and this paper, and a highlight of the pros and cons of each idea. Our conclusion is different from the negative of Norberg (2010). We believe that the whole idea of forward transition rates in multi-state models is relevant to actuarial practice, and we provide a substantiating example. Which version of the forward transition rates you should use depends heavily on what you want to use it for. This in itself does not diminuate the power of the concept but it exposes the demand for a thoughtful analysis of it. This is what we provide here.

As mentioned, one area of application is life insurance where finite-state models are generally accepted as the fundamental tool for representation of payment streams and their expected (present) values. However, the idea of forward transition rates is also potentially applicable in credit risk theory. Many credit derivatives specify nominal payments upon transition of a firm's state of financial health to a different state of financial health. Other derivatives specify nominal payments if a firm's financial health is in a specific state in the future. These are exactly the payments also evaluated in life insurance, see also Kraft and Steffensen (2007).

A key difference between life insurance and credit risk theory is that for life insurance, the transition rates can often reasonably well be assumed to be independent of the interest rate. Uncertainty of a wide range of transition rates is mainly driven by socio-demographic developments which are presumably not, at least not by first order, linked to the economy as such. Thereby, the difficulties arising from correlation between interest and mortality rates in a survival model, pointed out by Miltersen and Persson (2005), are of little relevance within that domain. In credit risk theory, the uncertainty of the transition rates is mainly driven by socio-economic developments that are, in contrast, strongly linked to the development of the interest rates.

In this exposition, we pay only little attention to the interest rate. Our focus is not on handling correlation with the interest rate but on handling, in the case of no correlation with the interest rate, the challenges arising from generalizing from the survival model to multi-state models. Therefore, it is targeted users of multi-state models in general and those in the area of life insurance in particular. The generalization to include correlation with interest rates is outside the scope here and is, together with discussing the dynamics of forward transition rate curves, postponed to future research.

The article is structured in the following way. In Section 5.2 we present the probabilistic setup. In Section 5.3 we define our new concept of forward transition rates. In Section 5.4 we compare our definition with other suggestions in the literature, and we discuss how to partly 'repair' the lack of match with transition densities. In Section 5.5 we relate the work to actuarial practice. Section 5.6 concludes.

5.2 Setup and background

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some background probability space. Also, let $J \in \mathbb{N}$ and let $S = \{0, 1, \ldots, J\}$ be some finite state space. In what follows, we consider a doubly stochastic Markov setting where the state of the insured is described by a jump process (chain) with values in S. Instead of working with an abstract filtered probability space satisfying the usual conditions, we recall an explicit construction. Details can be found in Jacobsen (2006, Chapter 3 and Sections 7.1-7.2). The approach can be considered somewhat restrictive, but it allows a simpler and more concise discussion of forward transition rates.

Notation and conventions Let $(Z_t)_{t\geq 0}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some measurable space. We denote with $\mathbb{F}^Z := (\mathcal{F}^Z_{[0,t]})_{t\geq 0}$ the natural filtration generated by Z and to which it itself is adapted, i.e.

$$\mathcal{F}^Z_{[0,t]} := \sigma(Z_s : s \leq t), \qquad t \in [0, \infty).$$

Furthermore, we define

$$\mathcal{F}^Z_{\infty} := \sigma\left(\bigcup_{t\geq 0} \mathcal{F}^Z_{[0,t]}\right),$$

$$\mathcal{F}^Z_{(t,\infty)} := \sigma(Z_s : s > t)\,, \qquad t \in [0, \infty).$$

We interpret $\mathcal{F}^Z_{\infty}$ as all the information generated by Z and $\mathcal{F}^Z_{(t,\infty)}$ as the future information generated by Z (after time t). Note that

$$\mathcal{F}^Z_{\infty} = \mathcal{F}^Z_{[0,t]} \vee \mathcal{F}^Z_{(t,\infty)},$$

where $\mathcal{F}^Z_{[0,t]} \vee \mathcal{F}^Z_{(t,\infty)} = \sigma(\mathcal{F}^Z_{[0,t]} \cup \mathcal{F}^Z_{(t,\infty)})$ is the join of $\mathcal{F}^Z_{[0,t]}$ and $\mathcal{F}^Z_{(t,\infty)}$.

In what follows, unless explicitly stated, all identities hold in an almost everywhere manner w.r.t. the probability measure $\mathbb{P}$.

Let $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{H}$ be sub-σ-algebras. We say that $\mathcal{G}_1$ and $\mathcal{G}_2$ are *conditionally independent given* $\mathcal{H}$ if

$$\mathbb{P}[\,G_1 \cap G_2 \,|\, \mathcal{H}] = \mathbb{P}[\,G_1 \,|\, \mathcal{H}]\, \mathbb{P}[\,G_2 \,|\, \mathcal{H}], \qquad \forall G_1 \in \mathcal{G}_1, G_2 \in \mathcal{G}_2. \tag{5.2.1}$$

We write $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{H}$ whenever (5.2.1) is satisfied. Finally, recall that (5.2.1) is equivalent to the asymmetric formulation

$$\mathbb{P}[G_1 \,|\, \mathcal{G}_2 \vee \mathcal{H}] = \mathbb{P}[G_1 \,|\, \mathcal{H}], \qquad \forall G_1 \in \mathcal{G}_1,$$

see e.g. Kallenberg (1997, Proposition 5.6).

5.2.1 Doubly stochastic Markov setting

For each possible transition $j, k \in S$, $k \neq j$, consider a stochastic process $[0,\infty) \ni t \mapsto \mu_{jk}(t)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $[0,\infty)$ and continuous sample paths. Using the Ionescu-Tulcea Theorem and the approach of Jacobsen (2006, Chapter 3 and Section 7.2), we can construct a jump process $X := (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in S, which has a deterministic initial state $x_0 \in S$ and, conditionally on μ, is Markovian with transition intensities μ. Here we take $(\Omega, \mathcal{F}, \mathbb{P})$ to be the canonical probability space associated with the construction.

That X is Markovian conditionally on μ means that for all $t \in [0,\infty)$,

$$\mathcal{F}^X_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^X_{[0,t]} \,\big|\, \sigma(X_t) \vee \mathcal{F}^\mu_\infty,$$

By construction, for $j, k \in S$, $k \neq j$, there exists (conditional) transition probabilities P^μ_{jk} such that for $0 \leq t < T < \infty$,

$$\mathbb{P}\Big[X_T = k \,\Big|\, \mathcal{F}^X_{[0,t]} \vee \mathcal{F}^\mu_\infty\Big] = P^\mu_{X_t k}(t,T).$$

Thus what we mean by the statement 'X conditionally on μ has transition intensities μ', is that (with $k \neq j$)

$$\lim_{h \searrow 0} \frac{1}{h} P^\mu_{jk}(t, t+h) = \mu_{jk}(t),$$

for all $t \in [0,\infty)$, which is well-defined as each μ_{jk} has continuous sample paths. Furthermore, the conditional transition probabilities P^μ_{jk} satisfy the Chapman-Kolmogorov equations and the *backward and forward integral and differential equations* (the so-called Feller-Kolmogorov equations).

In the following, we assume that $\mathbb{E}[\mu_{jk}(t)] < \infty$ for all $j, k \in S$, $k \neq j$, and all $t \in [0,\infty)$.

5.2.2 Preliminaries

In general, X is not unconditionally Markovian. An exception is whenever μ is deterministic; then X is trivially Markovian with transition intensities μ, and we recover the *classic Markov chain life insurance setting*, see e.g. Hoem (1969) and Norberg (1991).

From now on fix a time-point $t \in [0, \infty)$. For valuation of future liabilities and pricing in pension and life insurance, interest lies in the expected accumulated cash flow, in particular expressions of the form $\mathbb{E}[Z \mid \mathcal{F}^{X,\mu}_{[0,t]}]$, where Z is some $\mathcal{F}^{X,\mu}_{(t,\infty)}$/Borel-measurable random variable with values in $\mathbb{R}$ and finite expectation, $\mathbb{E}[|Z|] < \infty$. We think of Z as a future payment. In this paper, we disregard the time value of money and market risks and focus exclusively on the expected accumulated cash flow. If the market risks are assumed to be independent of the biometric and behavioral risks, the following results and discussions immediately extend to valuation taking the time value of money into account. Details are given in Section 5.6, where dependency between market risks and biometric and behavioral risks is also briefly discussed.

Because X is conditionally Markovian, we have the following result:

Lemma 5.2.1. *It holds that*

$$\mathcal{F}^{\mu}_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^{X}_{[0,t]} \,\big|\, \mathcal{F}^{\mu}_{[0,t]},$$
$$\mathcal{F}^{X,\mu}_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^{X}_{[0,t]} \,\big|\, \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}.$$

Proof. See Appendix 5.A. □

As an immediate consequence of the lemma, when Z is $\mathcal{F}^{X,\mu}_{(t,\infty)}$/Borel-measurable with values in $\mathbb{R}$ and $\mathbb{E}[|Z|] < \infty$, then

$$\mathbb{E}\Big[Z \,\Big|\, \mathcal{F}^{X,\mu}_{[0,t]}\Big] = \mathbb{E}\Big[Z \,\Big|\, \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}\Big]. \tag{5.2.2}$$

If Z furthermore is $\mathcal{F}^{\mu}_{(t,\infty)}$-measurable, then

$$\mathbb{E}\Big[Z \,\Big|\, \mathcal{F}^{X,\mu}_{[0,t]}\Big] = \mathbb{E}\Big[Z \,\Big|\, \mathcal{F}^{\mu}_{[0,t]}\Big]. \tag{5.2.3}$$

Therefore, we are really interested in quantities in the form

$$\mathbb{E}\Big[Z \,\Big|\, \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}\Big]$$

or simply

$$\mathbb{E}\Big[Z \,\Big|\, \mathcal{F}^{\mu}_{[0,t]}\Big].$$

These quantities are by definition either $\sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}$-measurable or simply $\mathcal{F}^{\mu}_{[0,t]}$-measurable, and we can therefore think of them as functions of the hitherto observed transition rates, and, if Z is only $\mathcal{F}^{X,\mu}_{(t,\infty)}$-measurable, also of the current state of the insured.

In the following, we shall use short hand notations such as $\mathcal{F}^{\mu}_{t}$ for $\mathcal{F}^{\mu}_{[0,t]}$, $\mathcal{F}^{X}_{t}$ for $\mathcal{F}^{X}_{[0,t]}$, and $\mathcal{F}^{X,\mu}_{t}$ for $\mathcal{F}^{X,\mu}_{[0,t]}$, etc.

5.2.3 Forward mortality

The concept of forward transition rates, which we introduce in the following section, is derived from the concept of forward mortality, which again is inspired by the concept of forward interest rates – for details we refer to Norberg (2010, Sections 2-4). To motivate the discussion on forward transition rates, we now recall the concept of forward mortality.

Consider a jump process X with values in $\{0,1\}$, which conditionally on μ is Markovian with transition intensities μ_{01} and $\mu_{10} = 0$. This setting corresponds to a survival model with stochastic mortality μ_{01}, see also Figure 5.1. Various authors, including Milevsky and Promislow (2001), Dahl (2004), and Dahl and Møller (2006), now essentially define the forward mortality rate as the $\mathcal{F}_t^\mu$-measurable and non-negative solution $(t,\infty) \ni T \mapsto m_{01}(t,T)$ to

$$\mathbb{E}\Big[e^{-\int_{(t,T]} \mu_{01}(s)\,\mathrm{d}s} \,\Big|\, \mathcal{F}_t^\mu\Big] = e^{-\int_{(t,T]} m_{01}(t,s)\,\mathrm{d}s}. \tag{5.2.4}$$

In what follows, we call m_{01} defined by (5.2.4) the *marginal forward mortality (rate)*, a choice of lingo which will become clear as we turn to the discussion of forward transition rate concepts in general.

Note that if we are allowed to interchange differentiation w.r.t. T and integration w.r.t. $\mathbb{P}$ in (5.2.4), the expression is equivalent to

$$\mathbb{E}\Big[e^{-\int_{(t,T]} \mu_{01}(s)\,\mathrm{d}s}\mu_{01}(T) \,\Big|\, \mathcal{F}_t^\mu\Big] = e^{-\int_{(t,T]} m_{01}(t,s)\,\mathrm{d}s} m_{01}(t,T). \tag{5.2.5}$$

Consider now simple accumulated payments $[0,\infty) \ni s \mapsto B(s)$ given by

$$\begin{aligned} \mathrm{d}B(s) &= \mathbb{1}_{\{X_s=0\}} b_0(s)\,\mathrm{d}s + b_{01}(s)\,\mathrm{d}X_s, \qquad s \in (0,\infty), \\ B(0) &= 0, \end{aligned}$$

where b_0 and b_{01} are continuous and real-valued deterministic functions. For valuation of future liabilities and pricing in pension and life insurance, interest lies in the expected accumulated cash flow, see e.g. Buchardt and Møller (2015). A general definition suitable for our setup follows below; here $\mathbb{H} = (\mathcal{H}_t)_{t\geq 0}$ refers to some filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ containing all relevant information accessible to the valuator, and B are accumulated payments assumed to be of finite variation, càdlàg and suitably integrable.

Definition 5.2.2. *Given information $\mathbb{H}$, the* ***expected accumulated cash flow*** *valuated at time $t \in [0,\infty)$ associated with the accumulated payments B is defined by*

$$(t,\infty) \ni T \mapsto A(t,T) := \mathbb{E}[B(T) - B(t) \,|\, \mathcal{H}_t].$$

$$\begin{tikzcd} \boxed{\text{alive} \quad 0} \arrow[r, "\mu_{01}(\cdot)"] & \boxed{\text{dead} \quad 1} \end{tikzcd}$$

Figure 5.1: *Survival model with stochastic mortality μ_{01}.*

We see that the expected accumulated cash flow valuated at time t with relevant filtration $\mathbb{H} = \mathbb{F}^{X,\mu}$ is given by

$$\begin{aligned} A(t,T) &= \mathbb{E}\Big[\mathbb{E}\big[B(T) - B(t) \,\big|\, \mathcal{F}_t^X \vee \mathcal{F}_\infty^\mu\big] \,\Big|\, \mathcal{F}_t^{X,\mu}\Big] \\ &= \mathbb{1}_{\{X_t=0\}} \int_{(t,T]} \mathbb{E}\Big[e^{-\int_{(t,s]} \mu_{01}(\tau)\,\mathrm{d}\tau} \left(b_0(s) + \mu_{01}(s)b_{01}(s)\right) \,\Big|\, \mathcal{F}_t^\mu\Big] \,\mathrm{d}s \\ &= \mathbb{1}_{\{X_t=0\}} \int_{(t,T]} e^{-\int_{(t,s]} m_{01}(t,\tau)\,\mathrm{d}\tau} \left(b_0(s) + m_{01}(t,s)b_{01}(s)\right) \mathrm{d}s, \end{aligned} \tag{5.2.6}$$

where we have used the tower property, that X conditionally on μ is Markovian with transition intensities μ_{01} and $\mu_{10} = 0$, equation (5.2.3), and that m_{01} satisfies (5.2.4) and (5.2.5).

When μ is deterministic, we recover the classic setting and the expected accumulated cash flow reads

$$\mathbb{1}_{\{X_t=0\}} \int_{(t,T]} e^{-\int_{(t,s]} \mu_{01}(\tau)\,\mathrm{d}\tau} \left(b_0(s) + \mu_{01}(s)b_{01}(s)\right) \mathrm{d}s. \tag{5.2.7}$$

Comparing (5.2.6) to (5.2.7) reveals exactly the prowess of the marginal forward mortality: It allows one to calculate the expected accumulated cash flow in the usual manner regardless of the fact that the mortality is stochastic by replacing the stochastic mortality with the marginal forward mortality in standard formulae. The wish for similar results for multi-state models motivates the concept of forward transition rates, which we study in the following section.

5.3 Forward equations rates

In this section, we first provide a detailed exposition on the concept of forward transition rates for the doubly stochastic Markov setting. In particular, we present the key properties which forward transition rate candidates desirably should satisfy. Motivated by this exposition, we next introduce the novel concept of *forward equations rates* and hereby provide new insights regarding the possibility of generalizing the concept of forward mortality for multi-state models. In the next section, we compare the forward equations rates to previous forward transition rate definitions in the literature. This is done in both an abstract manner and through a detailed example for disability insurance.

5.3.1 Forward transition rates

A natural question, as highlighted by Norberg (2010), is whether the concept of forward mortality can be adapted to and made fruitful in multi-state models, in particular the doubly stochastic Markov setting, or whether the results surveyed in the previous paragraph rely on the specific structure of the survival model, in which case a generalization is unobtainable. In addition to the work of Norberg (2010), the question has also been investigated by e.g. Christiansen and Niemeyer (2015) and Buchardt (2017).

To be more specific, the main question is if one can obtain similar replacement results regarding valuation as in the survival model. The main quantities of interest are

$$\begin{aligned}&\mathbb{1}_{\{X_T=k\}},\\&\mathbb{1}_{\{X_{T-}=k\}}\mu_{k\ell}(T),\end{aligned}$$

where $T \in (t,\infty)$ and $k,\ell \in S, \ell \neq k$. Why these quantities? Consider e.g. simple accumulated payments $[0,\infty) \ni s \mapsto B(s)$ given by

$$\begin{aligned}\mathrm{d}B(s) &= b_{X_s}(s)\,\mathrm{d}s + \sum_{\ell\in S}\mathbb{1}_{\{X_{s-}\neq\ell\}}b_{X_{s-}\ell}(s)\,\mathrm{d}N_{X_{s-}\ell}(s), \qquad s\in(0,\infty),\\ B(0) &= 0,\end{aligned}$$

where $N_{k\ell}$, $k,\ell\in S$, $\ell\neq k$, is the counting process counting the number of transitions from k to ℓ for X, and b_k and $b_{k\ell}$ are continuous real-valued deterministic functions describing the sojourn payments and payments upon transition, respectively. Omitting some primarily technical details, it follows that $[0,\infty)\ni s\mapsto M(s)$ given by $M(0)=0$ and

$$M(s) := B(s) - \int_{(0,s]}\Bigg(\sum_{k\in S}\mathbb{1}_{\{X_u=k\}}b_k(u) + \sum_{k,\ell\in S,\ell\neq k}\mathbb{1}_{\{X_{u-}=k\}}\mu_{k\ell}(u)b_{k\ell}(u)\Bigg)\mathrm{d}u$$

is a martingale w.r.t. the filtration $\mathbb{F}^X \vee \mathcal{F}^\mu_\infty$. By definition of the expected aggregated cash flow, it immediately becomes apparent why we are (solely, particularly) interested in the quantities $\mathbb{1}_{\{X_T=k\}}$ and $\mathbb{1}_{\{X_{T-}=k\}}\mu_{k\ell}(T)$.

Let $(t,\infty)\ni T\mapsto m_{jk}(t,T)$, $j,k\in S$, $k\neq j$, be $\sigma(X_t)\vee\mathcal{F}^\mu_t$-measurable candidate forward transition rates. To fully generalize the replacement argument obtained in the survival model, one needs that there exists differentiable $\sigma(X_t)\vee\mathcal{F}^\mu_t$-measurable functions $[t,\infty)\ni T\mapsto P^m_{X_tk}(t,T)$, satisfying

$$\begin{aligned}\frac{\partial}{\partial T}P^m_{X_tk}(t,T) &= \sum_{\ell\neq k}P^m_{X_t\ell}(t,T)m_{\ell k}(t,T) - P^m_{X_tk}(t,T)\sum_{\ell\neq k}m_{k\ell}(t,T), \qquad k\neq X_t\\ \sum_{k\in S}P^m_{X_tk}(t,T) &= 1, \\ P^m_{X_tk}(t,t) &= \mathbb{1}_{\{X_t=k\}}, \qquad k\in S,\end{aligned} \tag{5.3.1}$$

comparable to the Kolmogorov forward equations, such that

$$P^m_{X_tk}(t,T) = \mathbb{E}\big[\mathbb{1}_{\{X_T=k\}}\,\big|\,\sigma(X_t)\vee\mathcal{F}^\mu_t\big], \tag{5.3.2}$$

$$P^m_{X_tk}(t,T)m_{k\ell}(t,T) = \mathbb{E}\big[\mathbb{1}_{\{X_T=k\}}\mu_{k\ell}(T)\,\big|\,\sigma(X_t)\vee\mathcal{F}^\mu_t\big], \tag{5.3.3}$$

hold for all $k, \ell \in S$, $\ell \neq k$. Clearly, this boils down to two statements, that (5.3.1) and (5.3.2) hold or that (5.3.1) and (5.3.3) hold, or one stronger combined statement, namely that (5.3.1), (5.3.2), and (5.3.3) hold simultaneously. When referring to the first statement, we will often simply refer to (5.3.2) on its own. In similar fashion, we also do not explicitly mention (5.3.1) when referring to the second or the combined statement.

The identities (5.3.2) and (5.3.3) are the cornerstones for our approach due the following reason. When (5.3.2) holds, we have obtained successful replacement regarding the transition probabilities, while when (5.3.3) holds, we have obtained successful replacement regarding the transition densities. Thus when they hold simultaneously, the expected accumulated cash flow is given by

$$\begin{aligned} A(t,T) &= \mathbb{E}\Big[B(T) - B(t) \,\Big|\, \mathcal{F}_t^{X,\mu}\Big] \\ &= \int_{(t,T]} \sum_{k \in S} P^m_{X_t k}(t,s) \Big(b_k(s) + \sum_{\ell \in S, \ell \neq k} m_{k\ell}(t,s) b_{k\ell}(s) \Big) \, \mathrm{d}s, \end{aligned} \tag{5.3.4}$$

where we have employed similar techniques as in (5.2.6), that M is a martingale, and continuity of the (conditional) transition probabilities. The expected accumulated cash flow is then essentially in the 'usual form' known from the classic Markov chain life insurance setting, the only difference being that the stochastic transition intensities have been replaced by the forward transition rates – and thus a successful generalization of the replacement argument obtained in the survival model has been obtained.

Whenever m is non-negative and continuous, one can think of $P^m_{X_t k}(t, \cdot)$ as transition probabilities for a Markovian jump process with initial state X_t and transition intensities m conditionally on all the information up until and including time t, compare to the construction of X in the beginning of Subsection 5.2.1. As the forward transition rates m in general are only $\sigma(X_t) \vee \mathcal{F}_t^\mu$-measurable, the transition intensities and thus also the transition probabilities for this Markovian jump process depend on the current state X_t.

Note that we have yet to discuss existence and/or uniqueness of forward transition rate candidates satisfying (5.3.1), (5.3.2), and (5.3.3). The identities just represent desirable properties for any definition of forward transition rates. In the next subsection, we introduce a novel forward transition rate candidate based directly on (5.3.1) and (5.3.2).

5.3.2 Forward equations rates

In the following, we introduce a novel concept of *forward equations rates* and discuss existence, uniqueness and other properties regarding this forward transition rate definition.

Define the auxiliary $\mathcal{F}_t^\mu$-measurable function $[t,\infty) \ni T \mapsto \mathcal{P}_{jk}(t,T)$ by

$$\mathcal{P}_{jk}(t,T) = \mathbb{E}\Big[P_{jk}^\mu(t,T) \,\Big|\, \mathcal{F}_t^\mu \Big]$$

for each $j,k \in S$. **We assume in the following** that $\mathcal{P}_{jk}(t,\cdot)$ is differentiable for all $j,k \in S$. We then have the following forward transition rate definition.

Definition 5.3.1. *Let $(t,\infty) \ni T \mapsto m_{jk}(t,T)$, $j,k \in S$, $k \neq j$, be $\mathcal{F}_t^\mu$-measurable. If the following system of equations are satisfied,*

$$\begin{aligned}\frac{\partial}{\partial T}\mathcal{P}_{jk}(t,T) = &\sum_{\ell\neq k} \mathcal{P}_{j\ell}(t,T) m_{\ell k}(t,T) \\ &- \mathcal{P}_{jk}(t,T) \sum_{\ell \neq k} m_{k\ell}(t,T), \qquad j,k \in S, k \neq j,\end{aligned} \tag{5.3.5}$$

*we say that m are **forward equations rates** for X.*

This definition is similar to one suggested by Norberg (2010), yet there is a single but crucial difference. Norberg essentially suggests to define the forward transition rates as the $\sigma(X_t) \vee \mathcal{F}_t^\mu$-measurable solution to the system of equations

$$\begin{aligned}&\frac{\partial}{\partial T}\mathbb{E}\big[\mathbb{1}_{\{X_T = k\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu \big] \\ &\quad = \sum_{\ell \neq k} \mathbb{E}\big[\mathbb{1}_{\{X_T = \ell\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu \big]\, m_{\ell k}(t,T) \\ &\quad\quad - \mathbb{E}\big[\mathbb{1}_{\{X_T = k\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu \big] \sum_{\ell \neq k} m_{k\ell}(t,T), \qquad k \neq X_t,\end{aligned}$$

which is just (5.3.1) combined with (5.3.2). The definition imposed by (5.3.5) can be seen as an extension involving all transition probabilities rather than only those related to the present state of the insured, X_t. As such, (5.3.5) is a constrained version of Norberg's definition requiring the equations to hold for a portfolio of insured with different present states covering all states. As noted by Norberg, his system of equations consists of $J(J-1)$ unknowns but only $(J-1)$ equations, which in general would lead to infinitely many solutions. In comparison, (5.3.5) consists of $J(J-1)$ equations, so we actually expect the forward equations rates to exist and be unique (under suitable regularity conditions).

Together with the (trivially satisfied) conditions

$$\sum_{k\in S} \mathcal{P}_{jk}(t,T) = 1, \qquad \mathcal{P}_{jk}(t,t) = \mathbb{1}_{\{j=k\}}, \tag{5.3.6}$$

it can be shown that the forward equations rates are defined exactly such that (5.3.1) and (5.3.2) hold when setting $P_{X_t k}^m(t,\cdot) = \mathcal{P}_{jk}(t,\cdot)$ on $\{X_t = j\}$ for any $j \in S$ and all $k \in S$.

By definition, the forward equations rates are $\mathcal{F}_t^\mu$-measurable: They are not allowed to depend on the current state X_t. Later, when we investigate forward transition rate concepts in the literature, we return to a discussion of pros and cons regarding this property.

In general, the forward equations rates are allowed to be negative: In this case, one cannot think of $\mathcal{P}_{jk}(t,\cdot)$ as transition probabilities for some Markovian jump process, but must think of them as the solution to a system of differential equations similar to the Kolmogorov forward equations, namely (5.3.5) with conditions (5.3.6).

Regarding existence and uniqueness of the forward equations rates we have the following result, which is applicable for so-called decrement models, where return to a state is not possible once it has been left. Examples include the disability model without recovery with and without policyholder behavior (free policy and surrender).

Theorem 5.3.2. *Assume that $\mathcal{P}_{jk}(t,\cdot)$ is differentiable for all $j,k \in S$ and that $\mathcal{P}_{jk}(t,\cdot) = 0$ for $k < j$. Then the forward equations rates exist and are unique. If furthermore $\mathcal{P}_{jk}(t,\cdot)$ is continuously differentiable for all $j,k \in S$, then the forward equations rates are continuous.*

Proof. See Appendix 5.A. □

Regarding further properties of the forward equations rates, we note the following.

To conclude that $m_{jk}(t,\cdot) = 0$, one needs not only that direct transition from j to k is impossible but also that indirect transition from j to k is impossible. In other words, $\mu_{jk} = 0$ does not imply $m_{jk}(t,\cdot) = 0$ unless the stronger requirement $\mathcal{P}_{jk}(t,\cdot) = 0$ holds. This can be verified by e.g. considering a disability model without recovery where the mortality as active is zero. In particular, (5.3.3) does not hold in general. Rather, the closest obtainable identity involves the difference between the sum over all transitions to and from, respectively, each state. To be rigorous, it follows under the assumption of interchangeable differentiation w.r.t. T and integration w.r.t. $\mathbb{P}$, that

$$\begin{aligned}
&\sum_{\ell \neq k} P^m_{X_t \ell}(t,T) m_{\ell k}(t,T) - \sum_{\ell \neq k} P^m_{X_t k}(t,T) m_{k\ell}(t,T) \qquad (5.3.7)\\
=&\sum_{\ell \neq k} \mathbb{E}\big[\mathbb{1}_{\{X_T=\ell\}}\mu_{\ell k}(T) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu\big] - \sum_{\ell \neq k} \mathbb{E}\big[\mathbb{1}_{\{X_T=k\}}\mu_{k\ell}(T) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu\big],
\end{aligned}$$

for $k \in S$ using the definition of the forward equations rates m and the (conditional) Kolmogorov forward equations for X. This exactly shows that (5.3.3) holds only for the difference between the sum over transitions to and from, respectively, state k. For some specific classes of models this implies (5.3.3). To see this for competing risks models, note that for each death-state there is only one relevant transition, namely transition to this state from the alive-state, as the remaining transition intensities associated with the state are zero, whereby the above identity is identical to (5.3.3).

Because only (5.3.2) holds in general for the forward equations rates, the replacement argument of the survival model cannot be generalized fully. However, if the insurance contract does not contain payments upon transition, i.e. if $b_{k\ell} = 0$ for $k, \ell \in S$, $\ell \neq k$, then only (5.3.2) is required and the replacement argument generalizes. Thus if one is only interested in valuation of sojourn payments, the forward equations rates are a fruitful starting point.

Because the forward equations rates are defined directly from (5.3.2), the above discussion shows that any general definition of forward transition rates that is to satisfy both (5.3.2) and (5.3.3) cannot be $\mathcal{F}_t^\mu$-measurable (and thus must be allowed to depend on the current state X_t). This motivates the forward transition rate definition of Buchardt (2017) which we discuss in the following section.

5.4 Forward transition rate definitions in the literature

We now review the contributions of Christiansen and Niemeyer (2015) and Buchardt (2017) in comparison with the *forward equations rates*, which reveals strengths and weaknesses of each individual forward transition rate definition and leads to new insights regarding the forward transition rate concept in itself.

5.4.1 Alternative definitions from the literature

Christiansen & Niemeyer In Christiansen and Niemeyer (2015), forward transition rates are not discussed independently of the financial market, but this can easily be done by taking interest rate zero in their setting. Christiansen & Niemeyer define forward rates implicitly, essentially requiring they allow for replacement arguments comparable to ours for a set of insurance products. Based on the specific multi-state models and insurance products they consider, these requirements suggest the following definition: for each $j, k \in S$, $k \neq j$, the forward rate for this transition is the $\mathcal{F}_t^\mu$-measurable and non-negative solution $(t, \infty) \ni T \mapsto m_{jk}(t, T)$ to

$$\mathbb{E}\Big[e^{-\int_{(t,T]} \mu_{jk}(s)\,\mathrm{d}s} \,\Big|\, \mathcal{F}_t^\mu\Big] = e^{-\int_{(t,T]} m_{jk}(t,s)\,\mathrm{d}s}. \tag{5.4.1}$$

We note that m_{jk} does not depend on the current state of the insured: it only depends on the hitherto observed transition rates. Furthermore, the definition of m_{jk} does not involve the structure of the jump process X: they are 'universal'. In particular, the marginal forward mortality given by (5.2.4) is a special case of the general definition of (5.4.1). To see this, let X be a conditionally Markovian jump process given μ with values in $\{0, 1\}$ with $\mu_{10} = 0$. Then (5.4.1) is just (5.2.4) in disguise. Therefore, we call m_{jk} defined by (5.4.1) the *marginal forward transition* *rate*, as it solely relies on the probabilistic structure of μ. Consequentially, the marginal forward transition rates are particularly restrictive and idealistic.

Whenever the marginal forward transition rates satisfy (5.3.2), they agree with the forward equations rates. To see this, consider an active-surrender-dead model with transition rates from active to surrender and active to dead as in Figure 5.2. Then on $\{X_t = 0\}$ and with $k = 0$, we can restate (5.3.2) as

$$e^{-\int_{(t,T]}(m_{01}(t,s)+m_{02}(t,s))\,\mathrm{d}s} = \mathbb{E}\left[e^{-\int_{(t,T]}(\mu_{01}(s)+\mu_{02}(s))\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right]$$

On the other hand, by definition of the marginal forward transition rates, i.e. (5.4.1),

$$\begin{aligned}&e^{-\int_{(t,T]} m_{01}(t,s)\,\mathrm{d}s}e^{-\int_{(t,T]} m_{02}(t,s)\,\mathrm{d}s}\\ &\qquad= \mathbb{E}\left[e^{-\int_{(t,T]}\mu_{01}(s)\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right]\mathbb{E}\left[e^{-\int_{(t,T]}\mu_{02}(s)\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right].\end{aligned}$$

Collecting, we obtain the identity

$$\begin{aligned}&\mathbb{E}\left[e^{-\int_{(t,T]}(\mu_{01}(s)+\mu_{02}(s))\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right]\\ &\qquad= \mathbb{E}\left[e^{-\int_{(t,T]}\mu_{01}(s)\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right]\mathbb{E}\left[e^{-\int_{(t,T]}\mu_{02}(s)\,\mathrm{d}s}\,\middle|\,\mathcal{F}_t^{\mu}\right]\end{aligned}$$

which, unless μ_{01} and μ_{02} are independent, is not satisfied in general, see also Christiansen and Niemeyer (2015, Subsection 6.2) with interest rate zero. The situation is fully comparable to the discussion of forward mortalities and interest rates in the case of dependency between the biometric risks and the financial market, see e.g. Christiansen and Niemeyer (2015, Subsection 6.1), Miltersen and Persson (2005), and Buchardt (2014).

Christiansen and Niemeyer (2015) consider a large class of diffusion processes for the transitions rates μ and show the equivalence between specific dependency structures and the identities (5.3.2) and (5.3.3) for a number of multi-state models, including a disability model without recovery. Hereby, they show that if one desires 'universal' forward transitions rates that solely rely on the probabilistic structure of μ, such as the marginal forward transition rates, one must assume a specific and often unrealistic dependency structure between the transitions rates, see e.g. Christiansen and Niemeyer (2015, paragraph following Remark 5.5). If instead one is solely interested in sojourn payments and willing to specify a specific structure of the jump process X, the forward equations rates provide a natural alternative not confined to a specific dependency structure between the transition intensities.

Figure 5.2: *Active-surrender-dead model with transition rates μ_{01} and μ_{02}.*

Buchardt The definition of forward transition rates studied by Buchardt (2017) is for all practical purposes, see also Buchardt (2017, Lemma 4.3), equivalent to setting

$$(t,\infty) \ni T \mapsto m_{k\ell}(t,T) := \frac{\mathbb{E}\big[\mathbb{1}_{\{X_T=k\}}\mu_{k\ell}(T) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\mu}\big]}{\mathbb{E}\big[\mathbb{1}_{\{X_T=k\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\mu}\big]} \geq 0 \tag{5.4.2}$$

for all $k, \ell \in S$, $\ell \neq k$, whenever the right-hand side is well-defined. This definition was already proposed by Norberg (2010, Section 6, final paragraph).

We observe that if $\mu_{k\ell} = 0$, then $m_{k\ell}(t,\cdot) = 0$. Furthermore, from Buchardt (2017, Theorem 4.4) it follows that (5.3.2) holds (under some minor regularity conditions), such that by definition and rearrangement also (5.3.3) is satisfied. On the other hand, contrary to the forward equations rates and the marginal forward transition rates, the forward transition rates of (5.4.2) can by definition generally not be taken to be $\mathcal{F}_t^{\mu}$-measurable but must be allowed to depend on the current state X_t. Therefore, we call m_{kl} defined by (5.4.2) the *state-wise forward transition rate.*

In competing risks models, the state-wise forward transition rates agree with the forward equations rates (but in general differ from the marginal forward transitions rates unless the transition rates are assumed to be independent). If one imposes a specific structure on the transition intensities, this result can be extended beyond competing risks models – see also the example at the end of Section 5.5.

In the following subsection, similarities and differences between the state-wise forward transition rates and the forward equations rates are exemplified in the context of disability insurance. Furthermore, we exemplify how suitable state space and payment process 'tweaks' might allow for valuation of transition payments using the forward equations rates (by 'repairing' the lack of match with transition densities).

5.4.2 Disability insurance – 'repairing' the forward equations rates

Consider a disability model without recovery as in Figure 5.3. The insurance contract we have in mind is one stipulated by

- Premium payments when active, financing:
  - Disability coverage, including:
    - Payment upon transition from active to disabled.
    - Sojourn payments when disabled.
  - Death coverage given by payments upon transition to the state dead.

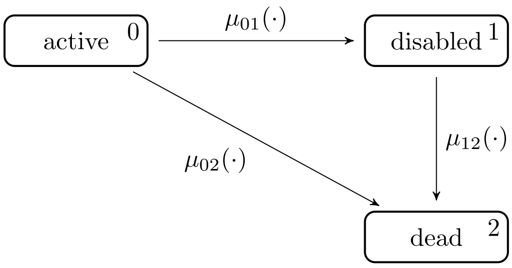

Figure 5.3: *Disability model without recovery.*

State-wise rates and forward equations rates Consider the state-wise forward transition rates from (5.4.2). These depend on the current state of the insured, thus we denote them by $m^{X_t}(t,\cdot)$. As discussed previously, both (5.3.2) and (5.3.3) are satisfied by the state-wise forward transition rates.

In general, $m_{12}^0(t,\cdot)$ and $m_{12}^1(t,\cdot)$ differ. On $\{X_t = 0\}$, i.e. when the insured is active at the present time, valuation of future sojourn payments and payments upon transition can be performed in a Markov model with $m^0(t,\cdot)$ as transition rates, see Figure 5.4 (left). On $\{X_t = 1\}$, i.e. when the insured is presently disabled, valuation must be performed in a Markov model with different transition rates $m^1(t,\cdot)$, see Figure 5.4 (right). In particular, four rather than three non-zero transition rates are required. From a practical and implementational point of view, valuation therefore remains slightly more complicated than in the classic Markov chain life insurance setting. Furthermore, the dependency of the forward transition rates on the current state of the insured makes them difficult to interpret.

Consider now instead the forward equations rates which we in a slight abuse of notation denote by $m(t,\cdot)$. As discussed previously, (5.3.3) is in general not satisfied by the forward equations rates – and this is also the case for the disability model without recovery. But valuation of future sojourn payments can be performed in a Markov model with $m(t,\cdot)$ as transition rates, see Figure 5.5. It can be shown that $m_{12}(t,\cdot) = m_{12}^1(t,\cdot)$ and that (5.3.3) does hold for the forward equations rates on

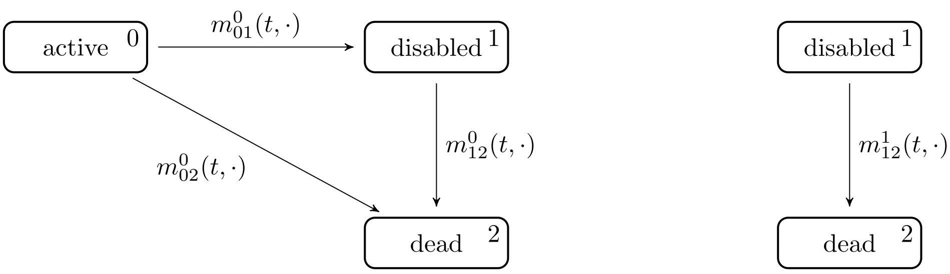

Figure 5.4: *Two Markov models with state-wise forward transition rates replacing the doubly stochastic Markov disability model without recovery: one to be used when the insured is presently active (left) and another to be used when the insured is presently disabled (right).*

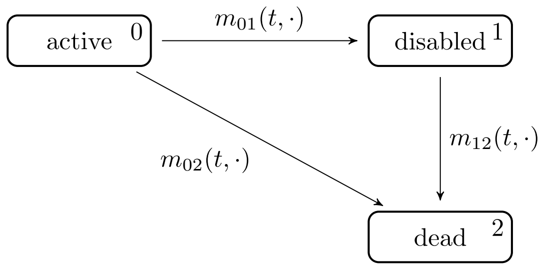

Figure 5.5: *Markov model with forward equations rates replacing the doubly stochastic Markov disability model without recovery for valuation of sojourn payments.*

$\{X_t = 1\}$, confer with (5.3.7). But in general, (5.3.3) does not hold on $\{X_t = 0\}$, in particular, the forward equations rates will not allow one to valuate transition payments from active to disabled or active to dead. To summarize, only parts of the original disability insurance contract we had in mind can be valuated if one insists on using forward equations rates. On the other hand, these parts – including the premiums payments when active and the sojourn payments when disabled – can be handled inside the technical and/or numerical framework of the classic Markov chain life insurance setting.

'Repairing' the forward equations rates We end this subsection by describing a way to tweak the original model slightly that extends the area of applicability of the forward equations rates.

Consider a new jump process $\tilde{X}$ defined from X by adding separate death states as in Figure 5.6. For all practical purposes, the models with and without separate death states are interchangeable as long as the sojourn payments in the two death states do not differ. To be rigorous, the new jump process satisfies

$$\tilde{X}_t = \mathbb{1}_{\{X_t \in \{0,1\}\}} X_t + \mathbb{1}_{\{N_{12}(t)=1\}} 2 + \mathbb{1}_{\{N_{02}(t)=1\}} 3,$$

where N is the multivariate counting process associated with X. Define $\tilde{\mu}$ as the corresponding (conditional) transition intensities, such that e.g. $\tilde{\mu}_{03} = \mu_{02}$, $\tilde{\mu}_{02} = 0$, and $\tilde{\mu}_{12} = \mu_{12}$. Then $\tilde{X}$ is also conditionally Markovian given $\tilde{\mu}$, as described

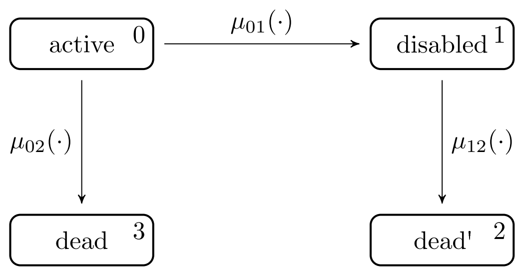

Figure 5.6: *Disability model without recovery but with separate death states.*

initially in Subsection 5.2.1, but contains a separate death state for death after disability.

One can show that while the state-wise forward transition rates remain unaffected, the forward equations rates for X and $\tilde{X}$ differ. Denote the latter by $\tilde{m}(t,\cdot)$. We cannot in general conclude that the forward equations rate $\tilde{m}_{02}(t,\cdot)$ is zero because indirect transition from state 0 to state 2 remains possible, which leads to the Markov model of Figure 5.7. In general, (5.3.3) does not hold on $\{X_t = 0\}$. Though when also $k = 0$ and $\ell = 3$, corresponding to valuation of payments upon transition from active to dead, (5.3.3) is satisfied, see e.g. (5.3.7). For valuation of payments upon transition from disabled to dead when the insured is presently active, we can rewrite (5.3.7) and obtain the following on $\{X_t = 0\}$:

$$\begin{aligned}\mathbb{E}\big[\mathbb{1}_{\{X_T=1\}}\mu_{12}(T)\,\big|\,\sigma(X_t)\vee\mathcal{F}_t^{\mu}\big] &= \mathbb{E}\Big[\mathbb{1}_{\{\tilde{X}_T=1\}}\tilde{\mu}_{12}(T)\,\Big|\,\sigma(\tilde{X}_t)\vee\mathcal{F}_t^{\tilde{\mu}}\Big] \\ &= \tilde{P}^{\tilde{m}}_{\tilde{X}_t0}(t,T)\tilde{m}_{02}(t,T) + \tilde{P}^{\tilde{m}}_{\tilde{X}_t1}(t,T)\tilde{m}_{12}(t,T).\end{aligned}$$

Thus for accumulated payments given by

$$\begin{aligned}\mathrm{d}B^{(1)}(s) &= \mathbb{1}_{\{X_{s-}=1\}}b_{12}(s)\,\mathrm{d}N_{X_{s-}2}(s), \qquad s\in(0,\infty), \\ B^{(1)}(0) &= 0,\end{aligned}$$

corresponding exactly to payment b_{12} upon transition from disabled to dead, the expected accumulated cash flow can on $\{X_t = 0\}$ be written as

$$A^{(1)}(t,T) = \int_{(t,T]} \Big(\tilde{P}^{\tilde{m}}_{00}(t,s)\tilde{m}_{02}(t,s)b_{12}(s) + \tilde{P}^{\tilde{m}}_{01}(t,s)\tilde{m}_{12}(t,s)b_{12}(s)\Big)\,\mathrm{d}s.$$

Thus valuation of the payments given by $B^{(1)}$ can be performed in the Markov model of Figure 5.7 with $\tilde{m}(t,\cdot)$ as transition rates through valuation of a different payment process with payment b_{12} upon transition from disabled to dead' as well as payment b_{12} upon transition from active to dead'.

Similar arguments apply for the payments upon transition from active to disabled. Here valuation can also be performed in a Markov model with $\tilde{m}(t,\cdot)$ as transition

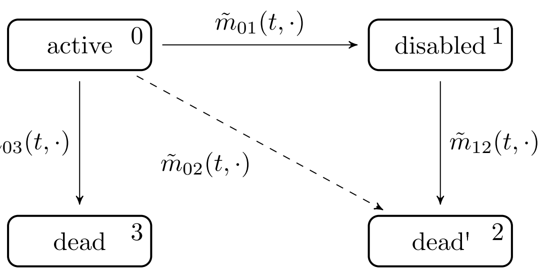

Figure 5.7: *Markov model with forward equations rates replacing the doubly stochastic Markov disability model without recovery with separate death states for alternative valuation of sojourn payments. Note the non-zero transition rate $\tilde{m}_{02}(t,\cdot)$ even though $\tilde{\mu}_{02}(\cdot) = 0$.*

rates through valuation of a different payment process with payment b_{01} upon transition from active to disabled as well as payment b_{01} upon transition from active to disabled-dead. Thus all parts of the original disability insurance contract we had in mind can be valuated in a Markov model, namely that of Figure 5.7, using forward equations rates if (and only if) one is willing to tweak the setup suitably. In particular, four rather than three non-zero transition rates are required. This means that from a practical and implementation point of view, valuation of payments upon transition remains slightly more complicated than in the classic Markov chain life insurance setting. Furthermore, the resulting forward transition rates are difficult to interpret.

Whether one works with state-wise forward transition rates or forward equations rates, we can conclude that four rather than three non-zero transition rates are required for the disability model without recovery. On the other hand, the above arguments do not generalize to arbitrary (non-decrement) models but rely extensively on the (decrement) structure of the disability model without recovery. So while it seems equally demanding to implement forward equations rates and state-wise forward transition rates (recall Figure 5.4) for valuation in the disability model without recovery, only implementation of the latter has a natural generalization to the most advanced models.

5.4.3 Summary and model calibration

All definitions discussed in the previous subsections extend the concept of forward mortality rates to a multi-state framework and contain the marginal forward mortality as a special case. The properties of the various forward transition rate definitions are summarized in Table 5.1. The extensions all express different ambitions. The definition of marginal forward transitions rates desires a sort of 'universality', in the sense that this definition does not rely on the specific structure of the state space or distribution of X but only relies relies on the probabilistic structure of μ. In general, this will not lead to a successful replacement argument, neither for sojourn payments, consult (5.3.2), nor payments upon transition, consult (5.3.3). In the definition of the forward equations rates, this condition is relaxed and only

	Universal	$\mathcal{F}_t^\mu$-measurable	(5.3.2)	(5.3.3)
Marginal rates	✓	✓		
Forward equations rates		✓	✓	
State-wise rates			✓	✓

Table 5.1: *Comparison of properties of the different forward transition rate definitions. Here the definition is said to be 'Universal' if it does not depend on the specific structure of the jump process but only relies on the probabilistic structure of the transition rates.*

$\mathcal{F}_t^\mu$-measurability is required, such that the rates still do not depend on the current state X_t of the insured. The replacement argument is then successful for sojourn payments but not in general for payments upon transition. Finally, the state-wise forward transition rates are allowed to depend on the current state of the insured, in which case the replacement argument holds for both sojourn payments and payments upon transition.

Another point of comparison between the definitions consists of comparing the quantities needed for a calibration similar to that of forward mortalities and forward interest rates.

To calibrate the marginal forward transition rates, we require the quantities

$$(t, \infty) \ni T \mapsto \mathbb{E}\Big[e^{-\int_{(t,T]} \mu_{jk}(s)\,\mathrm{d}s} \,\Big|\, \mathcal{F}_t^\mu\Big]$$

for $j, k \in S$, $k \neq j$. These quantities are not directly linked to any insurance contracts in the market.

To calibrate the forward equations rates, we require the quantities

$$(t, \infty) \ni T \mapsto \mathcal{P}_{jk}(t, T) = \mathbb{E}\Big[P_{jk}^\mu(t, T) \,\Big|\, \mathcal{F}_t^\mu\Big]$$

for $j, k \in S$. Assuming interest rate zero, these quantities are directly linked to insurance contracts consisting of sojourn payments.

To calibrate the state-wise forward transition rates, we require the quantities

$$\begin{aligned}
&(t, \infty) \ni T \mapsto \mathbb{E}\big[P_{X_t k}^\mu(t, T) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu\big], \\
&(t, \infty) \ni T \mapsto \mathbb{E}\big[P_{X_t k}^\mu(t, T)\mu_{k\ell}(T) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^\mu\big]
\end{aligned}$$

for $k, \ell \in S$, $\ell \neq k$. Assuming interest rate zero, these quantities are directly linked to insurance contracts consisting of sojourn payments and payments upon transition.

5.5 Forward-thinking and actuarial practice

Doubly stochastic extension of classic actuarial multi-state models allows for the inclusion of systematic (undiversifiable) risk and market consistent valuation in accordance with the Solvency II regulatory framework, see e.g. the discussion in the beginning of Buchardt (2014). In itself, multi-state modeling gives rise to computational complications, which historically have been circumvented by imposing a suitable Markovian structure, whereby the transition probabilities can be found by solving ordinary differential equations. In the classic Markov chain life insurance setting, the jump process describing the state of the insured is assumed Markovian, and the computational task is reduced to solving the system of Kolmogorov forward equations. This is not the case when considering doubly stochastic extensions, as any previous Markovian structure typically becomes void. In other words, the old weapons of the actuarial practitioner pose no threat to the new problems at hand. The development of mathematically sound definitions of forward transition rates is an attempt to once more stack the deck in favor of the actuarial practitioner. Conceptually, we are dealing with a whetstone for old weapons.

The practical relevance is essentially the following. The replacement conditions of (5.3.1)–(5.3.3) allow for a two-step valuation procedure: First, calibrate the forward transition rates, and then calculate the cash flow using classic numerical schemes (solving systems of ordinary differential equations). If the first step is not too demanding, the actuarial practitioner can avoid implementing new advanced numerical schemes and instead rely on already available platforms. This approach can be a valuable shortcut to the implementation of systematic risks in the practitioner's current valuation software.

Two-step procedures are of course not necessary; a general alternative is to solve the system of Kolmogorov forward partial integro-differential equations, see e.g. Buchardt (2017). But the two-step approach also shows its strengths in a conceptual sense: it transforms the computational complications to a question of calibration of forward transition rates. It is our belief that this transformation is beneficial to e.g. actuarial practitioners searching for simple benchmark models. A similar way of thinking in a slightly different framework drives the work of Christiansen and Niemeyer (2015). We provide a substantiating example at the end of this subsection.

The concept of forward transition rates is derived from the concept of forward mortality, which again is inspired by the concept of forward interest rates. In the context of the latter and as an alternative to short-rate modeling, Heath, Jarrow, and Morton (1992) propose a general framework, the so-called *Heath-Jarrow-Morton framework*, where the modeling object of interest is the entire forward interest rate curve. In the context of longevity risk, a similar framework has been developed by Bauer, Benth, and Kiesel (2012), where the marginal forward mortality curve rather than the stochastic mortality is the modeling object of interest. A similar change in modeling paradigm for multi-state settings might also prove valuable to practitioners. This requires mathematically sound definitions of forward transition rates (which we provide and discuss here) as well as the development of a framework similar to the Heath-Jarrow-Morton framework for doubly stochastic multi-state Markov models (which we have postponed to future research).

In the following example, we illustrate the relevance of forward transition rates to actuarial practice as discussed above, both from a conceptual as well as a computational point of view.

Survival model with surrender and free policy Consider the doubly stochastic model illustrated in Figure 5.8. We assume that η, ρ, ψ, and σ are non-negative and continuous, and that ψ and σ are also deterministic. Thus we allow for (possibly

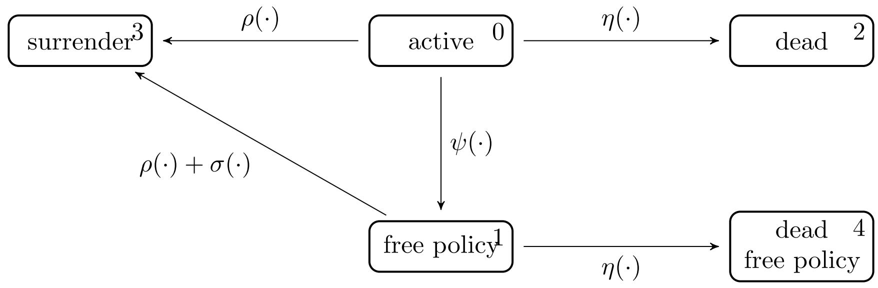

Figure 5.8: *Doubly stochastic survival model with options of surrender and conversion to free policy and with stochastic mortality and stochastic baseline surrender rate.*

dependent) stochastic mortality and stochastic surrender rates.

When η and ρ are deterministic and $\sigma = 0$, we are within the class of models considered by Buchardt and Møller (2015), see in particular Section 3.2 therein, where the connection to actuarial practice is also carefully explained.

Under certain regularity conditions, straightforward calculations (given in Appendix 5.A) show that the state-wise forward transition rates given by (5.4.2) take the form

$$
\begin{aligned}
m_{01}(t,T) &= \psi(T), \\
m_{02}(t,T) = m_{14}(t,T) &= \frac{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\eta(T)\,\Big|\,\mathcal{F}_t^{\eta,\rho}\Big]}{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\,\Big|\,\mathcal{F}_t^{\eta,\rho}\Big]}, \qquad (5.5.1)\\
m_{03}(t,T) &= \frac{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\rho(T)\,\Big|\,\mathcal{F}_t^{\eta,\rho}\Big]}{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\,\Big|\,\mathcal{F}_t^{\eta,\rho}\Big]}, \qquad (5.5.2)\\
m_{13}(t,T) &= m_{03}(t,T) + \sigma(T),
\end{aligned}
$$

with the state-wise forward transition rates being zero for the remaining indices.

The following two-step valuation procedure is now self-evident: First, calculate (5.5.1) and (5.5.2), and then calculate cash flows using classic methods. To illustrate the possible advantages of the two-step procedure within this example, assume that (η, ρ) belongs to the class of affine processes. Then (5.5.1) and (5.5.2) can be calculated by solving simple systems of ordinary differential equations, see e.g. Duffie, Pan, and Singleton (2000), Buchardt (2016), and Henriksen (2014, Chapter 5). In contrast, the general approach requires either solving the system of Kolmogorov forward partial integro-differential equations, see e.g. Buchardt (2017), or applying Monte Carlo methods. With reference to the study of numerical efficiency by Buchardt (2016) in a comparable setting, we conclude that in the affine setting, the two-step procedure is more efficient than the general approach.

In this example, the advantage of the two-step approach is illustrated using the state-wise forward transition rates, however, the conclusion also holds for the forward equations rates. On the basis of (5.5.1) and (5.5.2) there exists an $\mathcal{F}_t^{\eta,\rho}$-measurable version of the state-wise forward transition rates. In particular, under certain regularity conditions, the forward equations rates and the state-wise forward transition rates must agree, which implies that the forward equations rates satisfy (5.3.3). Note also that if η and ρ are independent, we obtain exactly the marginal forward transitions rates. It is the specific structure of the transition intensities within the model that makes the forward equations rates and the state-wise forward transition rates agree. Characterizing the class of models for which this is the case is postponed to future research

5.6 Concluding remarks

In the previous sections, we have focused solely on biometric and behavioral risks while not taking market risks and the time value of money into account. Hence we have only dealt with replacement arguments for the expected accumulated cash flow. In the context of reserving and pricing, interest lies in the prospective reserve, i.e. the expected present value of future payments. We now provide a short and informal discussion using market consistent valuation principles for life insurance and pensions, see e.g. Møller and Steffensen (2007, Chapter 3).

Let r be some continuous short rate. If the short rate is deterministic, then the prospective reserve V is simply given by

$$V(t) = \int_{(t,\infty)} e^{-\int_t^s r(u)\,\mathrm{d}u}\, A(t, \mathrm{d}s),$$

assuming the integral exists. If (5.3.2) and (5.3.3) hold, then

$$V(t) = \int_{(t,\infty)} e^{-\int_t^s r(u)\,\mathrm{d}u} \sum_{k\in S} P^m_{X_t k}(t,s)\Big(b_k(s) + \sum_{\ell\in S, \ell\neq k} m_{k\ell}(t,s) b_{k\ell}(s)\Big)\,\mathrm{d}s,$$

confer with (5.3.4). If the short rate is stochastic but the market risks are independent of the biometric and behavioral risks, the above instead reads

$$V(t) = \int_{(t,\infty)} e^{-\int_t^s f(t,u)\,\mathrm{d}u} \sum_{k\in S} P^m_{X_t k}(t,s)\Big(b_k(s) + \sum_{\ell\in S, \ell\neq k} m_{k\ell}(t,s) b_{k\ell}(s)\Big)\,\mathrm{d}s, \tag{5.6.1}$$

where $f(t,\cdot)$ is the usual forward interest rate associated with the short rate r. Thus as long as the markets risks are independent of the biometric and behavioural risks, the results and discussions of the previous sections extend from the expected accumulated cash flow to the prospective reserve in an immediate manner.

If there is dependency between the market risks and the biometric and behavioral risks, (5.6.1) ceases to hold and the previous results and discussions are not directly extendable. Forward transition and interest rates in the context of dependency between markets risks and biometric and behavioral risks are therefore not discussed in this paper. To our knowledge, only Buchardt (2014) has provided a forward rate concept allowing for successful replacement arguments in multi-state models with dependency between interest and transition rates. Buchardt (2014) only considers simple models consisting of at most one non-absorbing state. A natural next step is to extend the definition of forward equations rates and the definition of state-wise forward transition rates to allow for dependency between market risks and biometric and behavioral risks and compare the concepts to that of Buchardt (2014).

Acknowledgements

Christian Furrer's research is partly funded by the Innovation Fund Denmark under File No. 7038-00007B. We should like to thank Lars Frederik Brandt for fruitful discussions.

5.A Proofs

Before we prove Lemma 5.2.1, we recall the so-called chain rule for conditional independence, see e.g. Kallenberg (1997, Proposition 5.8). Let $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{K}$, and $\mathcal{H}$ be sub-σ-algebras. The chain rule states that

$$\mathcal{K} \perp\!\!\!\perp \mathcal{G}_1 \vee \mathcal{G}_2 \mid \mathcal{H}$$

if and only if

$$\mathcal{K} \perp\!\!\!\perp \mathcal{G}_1 \mid \mathcal{H} \quad \text{and} \quad \mathcal{K} \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{G}_1 \vee \mathcal{H}.$$

If $\mathcal{G}_1 \subset \mathcal{G}_2$ we find that

$$\mathcal{K} \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{H} \quad \Rightarrow \quad \mathcal{K} \perp\!\!\!\perp \mathcal{G}_1 \mid \mathcal{H},$$

which is called *reduction*, and

$$\mathcal{K} \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{H} \quad \Rightarrow \quad \mathcal{K} \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{G}_1 \vee \mathcal{H}. \tag{5.A.1}$$

Proof of Lemma 5.2.1. Because X is Markovian conditionally on μ, it holds that

$$\mathcal{F}^X_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^X_{[0,t]} \,\big|\, \sigma(X_t) \vee \mathcal{F}^\mu_\infty. \tag{5.A.2}$$

Furthermore, by construction, the conditional distribution of $(X_s)_{s \leq t}$ given $\mathcal{F}^\mu_\infty$ is $\mathcal{F}^\mu_{[0,t]}$-measurable: it is only a function of μ through $(\mu_s)_{s\leq t}$, confer with the properties of the (conditional) transition probabilities P^μ_{jk}. It follows from Kallenberg (1997, Proposition 5.6) that

$$\mathcal{F}^\mu_\infty \perp\!\!\!\perp \mathcal{F}^X_{[0,t]} \,\big|\, \mathcal{F}^\mu_{[0,t]},$$

Then by reduction,

$$\mathcal{F}^{\mu}_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^{X}_{[0,t]} \,\big|\, \mathcal{F}^{\mu}_{[0,t]}, \tag{5.A.3}$$

which proves the first part of the result. Beginning with (5.A.3), apply (5.A.1) to obtain

$$\mathcal{F}^{\mu}_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^{X}_{[0,t]} \,\big|\, \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}. \tag{5.A.4}$$

Let $\mathcal{K} = \mathcal{F}^{X}_{[0,t]}$, $\mathcal{H} = \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]}$, $\mathcal{G}_1 = \mathcal{F}^{\mu}_{(t,\infty)}$, and $\mathcal{G}_2 = \mathcal{F}^{X}_{(t,\infty)}$. Recall that $\mathcal{F}^{\mu}_{\infty} = \mathcal{F}^{\mu}_{[0,t]} \vee \mathcal{F}^{\mu}_{(t,\infty)}$. We might recast (5.A.4) and (5.A.2) as

$$\mathcal{K} \perp\!\!\!\perp \mathcal{G}_1 \mid \mathcal{H} \quad \text{and} \quad \mathcal{K} \perp\!\!\!\perp \mathcal{G}_2 \mid \mathcal{G}_1 \vee \mathcal{H},$$

and then it follows from the chain rule that

$$\mathcal{G}_1 \vee \mathcal{G}_2 \perp\!\!\!\perp \mathcal{K} \mid \mathcal{H}.$$

This exactly reads

$$\mathcal{F}^{X,\mu}_{(t,\infty)} \perp\!\!\!\perp \mathcal{F}^{X}_{[0,t]} \,\big|\, \sigma(X_t) \vee \mathcal{F}^{\mu}_{[0,t]},$$

which establishes the second part of the result completing the proof. □

Proof of Theorem 5.3.2. We first show that there exists a unique solution to (5.3.5) for any $T \in (t,\infty)$. Fix $T \in (t,\infty)$. In what follows we suppress t notationally.

With $m_{jj} := -\sum_{k \neq j} m_{jk}$, we can then rewrite (5.3.5) as

$$\frac{\partial}{\partial T}\mathcal{P}(T) = \mathcal{P}(T)m(T) \tag{5.A.5}$$

with $\frac{\partial}{\partial T}\mathcal{P}$, $\mathcal{P}$, and m being the corresponding matrices. Because $\mathcal{P}_{jk}(T) = 0$ for $k < j$, it follows that $\mathcal{P}(T)$ is an upper triangular matrix with diagonal elements $\mathcal{P}_{jj}(T)$. Now note that

$$P^{\mu}_{jj}(T) = \exp\left\{ -\int_t^T \sum_{k>j} \mu_{jk}(s)\,\mathrm{d}s \right\} > 0.$$

Consequently, it holds that $\mathcal{P}_{jj}(T) > 0$, hence in particular

$$\det \mathcal{P}(T) = \prod_{j=0}^{J} \mathcal{P}_{jj}(T) > 0.$$

This implies that $\mathcal{P}(T)$ is invertible with inverse $\mathcal{P}^{-1}(T)$, which is also an upper triangular matrix. Hence for fixed $T \in (t,\infty)$ there exists a unique solution given by

$$m(T) = \mathcal{P}^{-1}(T)\frac{\partial}{\partial T}\mathcal{P}(T).$$

Note that

$$
\begin{aligned}
\sum_{k \neq j} m_{jk}(T) &= \sum_{k \neq j} \sum_{\ell} \mathcal{P}_{j\ell}^{-1}(T) \frac{\partial}{\partial T} \mathcal{P}_{\ell k}(T) \\
&= \sum_{\ell} \mathcal{P}_{j\ell}^{-1}(T) \frac{\partial}{\partial T} \sum_{k \neq j} \mathcal{P}_{\ell k}(T) \\
&= - \sum_{\ell} \mathcal{P}_{j\ell}^{-1}(T) \frac{\partial}{\partial T} \mathcal{P}_{\ell j}(T) = -m_{jj}(T),
\end{aligned}
$$

according to the definition imposed in the beginning of the proof. This completes the proof of existence and uniqueness.

To complete the proof, we have to show that the solution is $\mathcal{F}_t^{\mu}$-measurable (as a function of T). This follows immediately by e.g. Cramer's rule because the entries of $\mathcal{P}$ are $\mathcal{F}_t^{\mu}$-measurable.

If $T \mapsto \frac{\partial}{\partial T} \mathcal{P}(t,T)$ is assumed to be continuous, it follows by similar arguments and an application of e.g. Cramer's rule that the solution also is continuous. $\square$

Forward transition rates in the survival model with surrender and free policy. We consider the state-wise forward transition rates given by (5.4.2). Because ψ and σ are deterministic, the only non-trivial derivations are related to m_{13} and m_{14}. By setting $\sigma = 0$ and using symmetry, the derivation of m_{14} follows from the derivation of m_{13}. Hence it suffices to derive m_{13}. On $\{X_t = 1\}$ it holds that

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{1}_{\{X_T=1\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\eta,\rho}\big] \\
&\qquad = \mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s)+\sigma(s))\,\mathrm{d}s} \,\Big|\, \mathcal{F}_t^{\eta,\rho}\Big], \\
&\mathbb{E}\big[\mathbb{1}_{\{X_T=1\}}\,(\rho(T)+\sigma(T)) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\eta,\rho}\big] \\
&\qquad = \mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s)+\sigma(s))\,\mathrm{d}s}(\rho(T)+\sigma(T)) \,\Big|\, \mathcal{F}_t^{\eta,\rho}\Big].
\end{aligned}
$$

Consequently,

$$
m_{13}(t,T) = \sigma(T) + \frac{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\rho(T) \,\Big|\, \mathcal{F}_t^{\eta,\rho}\Big]}{\mathbb{E}\Big[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s} \,\Big|\, \mathcal{F}_t^{\eta,\rho}\Big]}
$$

on $\{X_t = 1\}$. Let now C be defined by

$$
C(t,T) = \int_{(t,T]} e^{-\int_{(t,s]}\psi(u)\,\mathrm{d}u}\psi(s)e^{-\int_{(s,T]}\sigma(u)\,\mathrm{d}u}\,\mathrm{d}s.
$$

Note that on $\{X_t = 0\}$,

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{1}_{\{X_T=1\}} \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\eta,\rho}\big] \\
&= \mathbb{E}\left[\int_{(t,T]} e^{-\int_{(t,s]}(\eta(u)+\rho(u)+\psi(u))\,\mathrm{d}u}\psi(s)e^{-\int_{(s,T]}(\eta(u)+\rho(u)+\sigma(u))\,\mathrm{d}u}\,\mathrm{d}s \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right] \\
&= \mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s} \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right] C(t,T), \\
&\mathbb{E}\big[\mathbb{1}_{\{X_T=1\}}\,(\rho(T)+\sigma(T)) \,\big|\, \sigma(X_t) \vee \mathcal{F}_t^{\eta,\rho}\big] \\
&= \mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\,(\rho(T)+\sigma(T)) \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right] C(t,T).
\end{aligned}
$$

Thus whenever ψ is strictly positive on a subset of $(t,T]$ with non-zero Lebesgue measure, it holds on $\{X_t = 0\}$ that

$$
\begin{aligned}
m_{13}(t,T) &= \frac{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}(\rho(T)+\sigma(T)) \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]}{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s} \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]} \\
&= \sigma(T) + \frac{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\rho(T) \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]}{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s} \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]},
\end{aligned}
$$

as the terms involving $C(t,T)$ cancel. To conclude, this shows that

$$
m_{13}(t,T) = \sigma(T) + \frac{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s}\rho(T) \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]}{\mathbb{E}\left[e^{-\int_{(t,T]}(\eta(s)+\rho(s))\,\mathrm{d}s} \,\middle|\, \mathcal{F}_t^{\eta,\rho}\right]}
$$

is an $\mathcal{F}_t^{\eta,\rho}$-measurable version of the state-wise forward transition rates. $\square$

Chapter 6

Experience rating in the classic Markov chain life insurance setting: An empirical Bayes and multivariate frailty approach

This chapter contains the paper *Furrer (2019)*.

Abstract

We consider experience rating in the classic Markov chain life insurance setting. We focus on shrinkage estimation of group effects in an empirical Bayes and multivariate frailty extension, building on ideas from group life insurance and survival and event history analysis. Within this framework, we provide insights regarding the structure of the likelihoods and sufficiency of summary statistics such as occurrences and exposures. Simple shrinkage estimators, given by well-known credibility formulas, are obtained under quadratic loss for mutually independent conjugate Gamma priors. The applicability of these simple shrinkage estimators for disability insurance is illustrated in a numerical example using simulated data.

Keywords: Classic Markov chain life insurance setting; Empirical Bayes; Experience rating; Multivariate frailty; Shrinkage

6.1 Introduction

We consider experience rating in multi-state life insurance for groups of insured aiming to accurately assess group performances. The modern (stochastic) modeling approach in life insurance was introduced by Janssen (1966) and Hoem (1969) and is based on representing the biometric states (active, disabled, dead, etc.) of the insured by jump processes on a finite state space. Special attention has been paid to the *classic Markov chain life insurance setting*, which is obtained by considering independent Markovian jump processes, and its semi-Markovian extension, see e.g. Norberg (1991) and Christiansen (2012).

Experience rating in multi-state life insurance is non-trivial. Reserves and premiums are given by present values of stochastic payment streams generated by jump processes, which implies that most methods from (linear Bayes) credibility theory cannot be applied directly as these methods rely on explicit or implicit models for the risk premium itself. In particular, the classic Bühlmann-Straub model, see Bühlmann and Straub (1970), and its extensions fall short in their original formulation. In regards to group mortality modeling with a known population mortality table, suitable reformulations and subsequent applications of the Bühlmann-Straub model are given in Norberg (1989b, Sections 3-4) and Klugman et al. (2009).

Experience rating is essentially about (optimal) shrinkage estimation of group effects. In this paper, we propose a simple shrinkage estimation procedure for the classic Markov chain life insurance setting by introducing multivariate frailty and applying empirical Bayes methods. Hereby we extend the work of Norberg (1989b) for group life insurance to a multi-state setting. Our contribution consists of two parts. The first part relates specifically to the simple shrinkage estimation procedure. We identify the need for experience rating through shrinkage estimation in multi-state life insurance, we propose a simple shrinkage estimator, we interpret it as an empirical Bayes estimator in a multivariate frailty extension of the classic setting, and we illustrate the resulting estimation procedure by a numerical example for disability insurance using simulated data. The second part constitutes a careful analysis of the (product) structure of the marginal likelihood in the extended setting from which we establish sufficient conditions for the original inferential problem to split into a number of sub-problems. These results are of limited theoretical value, but they do provide the actuarial practitioner with another tool to identify simpler, and thus also more interpretable, classes of models.

The paper targets actuarial researchers and practitioners alike. Throughout the paper, we elaborate on the link between jump process modeling and the use of Poisson (mixture) regressions for summary statistics such as occurrences and exposures – building a bridge between theory and practice.

The article is structured as follows. In Section 6.2, we recall parametric modeling and estimation for the classic Markov chain life insurance setting. In Section 6.3, we extend the classic setting to allow for (latent) group heterogeneity. Based on the parametric structure and the (in)dependency between the latent variables, we give a characterization of the (product) structure of the likelihood from which we identify sufficient statistics and motivate a simple shrinkage estimation procedure. In Section 6.4, we provide a numerical example for disability insurance, illustrating the applicability of the simpler shrinkage estimation procedure. Section 6.5 concludes with a discussion on model extensions.

6.2 The classic Markov chain life insurance setting

In this section, we briefly recall parametric modeling and estimation for the classic Markov chain life insurance setting as inference for multivariate counting processes. The exposition follows along the lines of the standard text book reference Andersen et al. (1988).

Next, we associate, in an approximate manner, the relevant likelihoods of the classic Markov chain life insurance setting with those stemming from specific Poisson regressions, and we discuss sufficiency of summary statistics such as occurrences and exposures.

In practice and regarding e.g. group life and group disability insurance, it is crucial from the insurers point of view to be able to identify low- and high-risk groups using past data of said groups (experience rating) and develop pricing and reserving models taking these differences in risks into account to avoid serious issues arising from adverse selection phenomena. We outline a simple framework within the classic Markov chain life insurance setting in which experience rating relates to shrinkage estimation of group intercepts for the transition intensities. This framework will serve as the starting point for the multivariate frailty and empirical Bayes approach to experience rating that follows in Section 6.3.

6.2.1 Setup, parametric modeling, maximum likelihood estimation, and sufficiency

In the following, we consider a filtered probability space $(\Omega, \mathbb{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions except possibly completeness. Let $S = \{0, 1, \ldots, J\}$ for some $J \in \mathbb{N}$, and let $X = (X(t))_{t\geq 0}$ be a jump process on S with deterministic initial value $X(0) = x(0) \in S$.

We can interpret S as the relevant biometric and behavioral states of the insured, e.g. active, disabled, free-policy, and dead, whereby X describes the state of the insured through time.

Denote with $N = (N_{jk}(t))_{j,k\in S,k\neq j,t\geq 0}$ the multivariate counting process associated with X defined by $N_{jk}(0) = 0$ and

$$(0, \infty) \ni t \mapsto N_{jk}(t) = \#\{s \in (0, t] : X(s-) = j \text{ and } X(s) = k\}.$$

We assume that N_{jk} has intensity process λ_{jk} given by

$$(0, \infty) \ni t \mapsto \lambda_{jk}(t) = \mathbb{1}_{(X(t-)=j)}\alpha_{jk}(t)$$

for measurable, positive functions $[0,\infty) \ni t \mapsto \alpha_{jk}(t)$, assumed bounded on bounded intervals. In particular,

$$[0,\infty) \ni t \mapsto N_{jk}(t) - \int_0^t \lambda_{jk}(s)\,\mathrm{d}s$$

are martingales w.r.t. $\mathcal{F}$, and X is Markovian with *transition intensities* α. For the canonical construction of X and N using marked point processes, which in particular implies the existence of a probabilistic setting for which the aforementioned assumptions hold, see e.g. Jacobsen (1982, 2006).

In the context of life insurance, X is of course at best only observed in a finite time interval $[0,\tau]$ for some deterministic right-censoring time $\tau \in (0,\infty)$. Thus, X is seldomly fully observed – the exception being the occurrence of absorption, e.g. death in a disability model, before time τ.

In general, insured enter and exit the insurance portfolio randomly, which leads to general left-truncation and right-censoring phenomena. The complications arising hereby can be handled within the martingale theory for multivariate counting processes, but to keep the exposition from becoming unnecessarily technical, we restrict our attention to individual deterministic right-censoring. Finally, in Section 6.5, we briefly sketch out how to obtain similar results when allowing for general left-truncation and right-censoring.

Consider now a parametric model for the transition intensities where α is parametrized by a parameter vector β with values in some parameter set $\mathcal{B}$. To be precise, let $(\mathbb{P}_\beta)_{\beta\in\mathcal{B}}$ be a family of probability measures on $(\Omega,\mathbb{F},\mathcal{F})$, and assume that under $\mathbb{P}_\beta$, X has transition intensities $\alpha(t;\beta)$ in the sense that $N_{jk}(t)$ has intensity process $\lambda_{jk}(t;\beta) = \mathbb{1}_{(X(t-)=j)}\alpha_{jk}(t;\beta)$ under $\mathbb{P}_\beta$. The likelihood of $(N(t))_{t\in[0,\tau]}$, equivalently that of $(X(t))_{t\in[0,\tau]}$, is

$$\mathcal{L}(\beta) = \prod_{j,k\in S, k\neq j} \exp\left\{ \int_{(0,\tau]} \log\left(\alpha_{jk}(t;\beta)\right) \mathrm{d}N_{jk}(t) - \int_0^\tau \mathbb{1}_{(X(t-)=j)}\alpha_{jk}(t;\beta)\,\mathrm{d}t \right\}. \tag{6.2.1}$$

This as well as the following likelihoods are suitable Radon-Nikodyms derivatives of the restriction of the distribution of N, equivalently X, to $\mathbb{F}_\tau$ w.r.t. some σ-finite reference measure that does not depend on β.

The product structure of (6.2.1) w.r.t. the transitions is important when it comes to maximum likelihood estimation and identification of sufficient statistics. To see this, consider the extreme case where each transition is distinctly parametrized, i.e. let $\mathcal{B} = \mathcal{B}_{01} \times \mathcal{B}_{02} \times \cdots \times \mathcal{B}_{J(J-1)}$, and assume that $\alpha_{jk}(\cdot;\beta)$ only depends on β through β_{jk}. In this case (6.2.1) reads

$$\mathcal{L}(\beta) = \prod_{j,k\in S, k\neq j} \mathcal{L}_{jk}(\beta_{jk})$$

where $\mathcal{L}_{jk}$ is given by

$$\mathcal{L}_{jk}(\beta_{jk}) = \exp\left\{\int_{(0,\tau]} \log\left(\alpha_{jk}(t;\beta_{jk})\right) \mathrm{d}N_{jk}(t) - \int_0^\tau \mathbb{1}_{(X(t-)=j)}\alpha_{jk}(t;\beta_{jk})\,\mathrm{d}t\right\}.$$

Thus the maximum likelihood estimator of β_{jk} is simply the maximizer of $\mathcal{L}_{jk}(\beta_{jk})$, and each β_{jk} can be estimated independently from the remaining parameters. Furthermore, $(\mathbb{1}_{(X(t)=j)}, N_{jk}(t))_{t\in[0,\tau]}$ is a sufficient statistic for β_{jk}. All in all, the original inferential problem splits into $(J+1)J$ sub-problems where the (j,k)'th sub-problem only involves information related to the (j,k)'th transition given by the indicator of being in state j and the process counting the number of jumps from state j to state k. From a purely theoretical point of view, these split-ups might appear rather unimportant and trivial, but to an actuarial practitioner the results are essential as they allow him/her to take a marginal approach to data collection and risk estimation.

In the case where each transition is not distinctly parametrized but there still remains some structure in the parametrization of the transition intensities, similar results regarding maximum likelihood estimation and sufficiency as above can be obtained. Rather than carrying out a general discussion, we provide a relevant example regarding disability insurance from which one can easily extrapolate the general results:

Example 6.2.1. Let $S = \{0,1,2\}$ with 0 denoting active, 1 denoting disabled, and 2 denoting dead, and assume that $\alpha_{10} = \alpha_{20} = \alpha_{21} = 0$. The setting then corresponds to a classic Markovian disability model without recovery, see also Figure 6.1.

Consider now the case where the mortalities and disability rate are distinctly parametrized: let $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$, and assume that $\alpha_{jk}(\cdot)$ only depends on β through β_k. In this case, the mortalities are parametrized by β_2, while the disability rate is parametrized by β_1. Consequently, the likelihood $\mathcal{L}$ is the product of terms $\mathcal{L}_1$ and

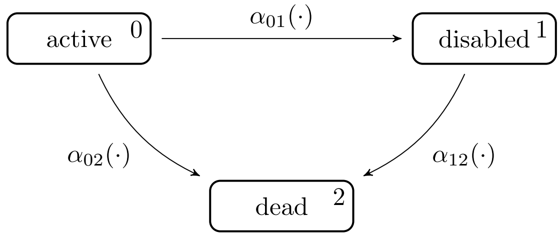

Figure 6.1: *Classic Markovian disability model without recovery.*

$\mathcal{L}_2$ given by

$$\begin{aligned}\mathcal{L}_2(\beta_2) = \exp&\left\{\int_{(0,\tau]} \log\left(\alpha_{02}(t;\beta_2)\right) \mathrm{d}N_{02}(t) - \int_0^\tau \mathbb{1}_{(X(t-)=0)}\alpha_{02}(t;\beta_2)\,\mathrm{d}t\right\}\\ \exp&\left\{\int_{(0,\tau]} \log\left(\alpha_{12}(t;\beta_2)\right) \mathrm{d}N_{12}(t) - \int_0^\tau \mathbb{1}_{(X(t-)=1)}\alpha_{12}(t;\beta_2)\,\mathrm{d}t\right\},\\ \mathcal{L}_1(\beta_1) = \exp&\left\{\int_{(0,\tau]} \log\left(\alpha_{01}(t;\beta_1)\right) \mathrm{d}N_{01}(t) - \int_0^\tau \mathbb{1}_{(X(t-)=0)}\alpha_{01}(t;\beta_1)\,\mathrm{d}t\right\}.\end{aligned}$$

It follows that the maximum likelihood estimator of β_k is the maximizer of the simpler term $\mathcal{L}_k$. Furthermore,

$$(\mathbb{1}_{(X(t)=0)}, N_{01}(t))_{t\in[0,\tau]}$$

is a sufficient statistic for β_1. Thus to estimate the disability rate, i.e. the parameters β_1, the necessary data is only the indicator of being active and the count of disabilities. On the other hand, we can only conclude from $\mathcal{L}_2$ that

$$(\mathbb{1}_{(X(t)=0)}, N_{02}(t), \mathbb{1}_{(X(t)=1)}, N_{12}(t))_{t\in[0,\tau]}$$

is a sufficient statistic for β_2. Observing this vector is equivalent to observing $(X(t))_{t\in[0,\tau]}$, so all data is still needed for estimation of β_2. This is a consequence of the specific state-space model and of the fact that α_{02} and α_{12} both depend on β_2. If the transitions $(0,2)$ and $(1,2)$ were also distinctly parametrized, data reduction would have taken place. ○

6.2.2 The link to Poisson regressions

Extending the parametric framework of the previous subsection, consider $h = 1,\dots,H$ independent Markov jump processes $X = (X^{(1)},\dots,X^{(H)})$ with associated multivariate counting processes $N = (N^{(1)},\dots,N^{(H)})$. Each insured is assumed observed up until and including time $\tau^{(h)} \in (0,\infty)$. Denote with τ the latest of these time-points, i.e. $\tau = \max_h \tau^{(h)}$. Let $z^{(h)} \in \mathcal{Z}$ (with $\mathcal{Z}$ finite) be a time-independent categorical covariate for insured h, denoting e.g. the year of birth and sex of the insured. Assume that $X^{(h)}$ has transition intensities $[0,\infty) \mapsto \alpha_{jk}(t;\beta|z^{(h)})$. Depending on the model, α_{jk} could e.g. be a Gompertz-Makeham mortality law with parameters depending on the sex of the insured.

In actuarial practice, estimation of the transition intensities in the above setting is often based on Poisson regressions or extensions thereof. Recent examples include the credibility model of Christiansen and Schinzinger applied to mortality data, Christiansen and Schinzinger (2016, Section 7), the application of machine learning techniques to mortality modeling, Deprez, Shevchenko, and Wüthrich (2017), and the benchmark model of the Danish FSA described in Jarner and Møller (2015,

Appendix 1). In the field of survival and event history analysis, it has long been known that this approach can be motivated by the structure of the likelihood, see e.g. Aalen, Borgan, and Gjessing (2008, Subsection 5.2.1). In the context of insurance, this was recently pointed out by Gschlössl, Schoenmaekers, and Denuit (2011). We recall the argument. Let $0 = t_0 < t_1 < \cdots < t_K = \tau$ be a grid of $[0, \tau]$, and assume that the transition intensity α_{jk} is continuous from the left and piecewise constant on this grid. This can in many cases be seen as an approximation to the true transition intensity with convergence as the grid becomes finer. The (j, k)'th term of the likelihood now reads

$$\prod_{i=1}^{K} \prod_{z \in \mathcal{Z}} \alpha_{jk}(t_i; \beta | z)^{O_{jk}(i|z)} e^{-E_j(i|z)\alpha_{jk}(t_i;\beta|z)} \tag{6.2.2}$$

where $O_{jk}(i|z)$ and $E_j(i|z)$ are given by

$$O_{jk}(i|z) = \sum_{h=1}^{H} \mathbb{1}_{(z^{(h)}=z)} \int_{(0,\tau^{(h)}]} \mathbb{1}_{(t_{i-1},t_i]} \, \mathrm{d}N_{jk}^{(h)}(t), \tag{6.2.3}$$

$$E_j(i|z) = \sum_{h=1}^{H} \mathbb{1}_{(z^{(h)}=z)} \int_{0}^{\tau^{(h)}} \mathbb{1}_{(t_{i-1},t_i]} \mathbb{1}_{(X^{(h)}(t-)=j)} \, \mathrm{d}t. \tag{6.2.4}$$

Thus $O_{jk}(i|z)$ is the number of *occurrences* of transition from state j to state k in time interval i for insured with covariate z, while $E_j(i|z)$ is the total time spent (*exposure*) in state j in time interval i for insured with covariate z.

The connection to Poisson regressions follows by the observation that (6.2.2) is proportional to the likelihood of independent observations

$$(O_{jk}(i|z))_{i=1,\dots,K; z \in \mathcal{Z}}$$

with distributions Poisson$(E_j(i|z)\alpha_{jk}(t_i; \beta|z))$ under $\mathbb{P}_\beta$, where the exposure $E_j(i|z)$ is considered fixed. Thus under the (approximate) assumption of piecewise constant transition intensities and for methods of inference that satisfy the likelihood principle, such as maximum likelihood estimation and empirical Bayes estimation, inference in the classic Markov chain life insurance setting can be based on Poisson regressions. Furthermore, the occurrences and exposures for each time interval and each possible covariate will be sufficient statistics. Therefore, the assumption of piecewise constant transition intensities allows one to build some kind of bridge between the classic Markov chain life insurance setting and actuarial practice, where often only summary statistics such as occurrences and exposures are available.

6.2.3 Estimation of group effects

Having clarified well-known but important facts regarding maximum likelihood estimation, sufficiency, and the relationship between the classic Markov chain life insurance setting and Poisson regressions, which play a major role for the investigation in Section 6.3, we now turn to the main topic of the paper.

Consider a modification of the setting of the previous Subsection 6.2.2 in which we assume that the insured are composed into $G \in \mathbb{N}$ independent groups of independent insured. We assume throughout the paper that the group composition is given a priori and is fixed in time, such as for group life insurance. An extension to time-dependent group compositions is finally briefly discussed in Section 6.5.

We write $N^{(g)} = (N^{(g1)}, \ldots, N^{(gH_g)})$ for the multivariate counting processes of insured in group g. Similar notation is used for the jump processes, covariates, and right-censoring times. Due to our interest in experience rating, we assume that the transition intensities depend on the group association and write $[0, \infty) \ni t \mapsto \alpha_{jk}(t|g, z^{(gh)})$ to emphasize this dependency. In this paper, we focus on models in the form

$$\alpha_{jk}(t; (\xi, \beta)|g, z^{(gh)}) = \xi_{jk}^{(g)} \, \mu_{jk}(t; \beta|z^{(gh)}),$$

where for each specific transition (j, k), μ_{jk} is a group-independent *base transition intensity* parametrized by β, while $\xi_{jk}^{(g)}$ is the (positive) effect for group g. The group effect has been introduced in such a way that it is constant (over time) and identical for insured in the same group. These assumptions can be sought to be relaxed, but that is not the focus of this paper. Focusing on a single transition (j, k) and (approximately) assuming μ_{jk} to be continuous from the left and piecewise constant on some grid $0 = t_0 < t_1 < \cdots < t_K = \max_{g,h} \tau^{(gh)}$, we know from the previous subsection that the (j, k)'th term of the likelihood reads

$$\prod_{g=1}^{G} \prod_{i=1}^{K} \prod_{z \in \mathcal{Z}} \left(\xi_{jk}^{(g)} \mu_{jk}(t_i; \beta|z) \right)^{O_{jk}(i|g,z)} e^{-E_j(i|g,z) \xi_{jk}^{(g)} \mu_{jk}(t_i;\beta|z)}$$

where $O_{jk}(i|g, z)$ and $E_j(i|g, z)$ are the groupwise occurrences and exposures, i.e.

$$O_{jk}(i|g, z) = \sum_{h=1}^{H_g} \mathbb{1}_{(z^{(gh)}=z)} \int_{(0,\tau^{(gh)}]} \mathbb{1}_{(t_{i-1},t_i]} \, \mathrm{d}N_{jk}^{(gh)}(t), \tag{6.2.5}$$

$$E_j(i|g, z) = \sum_{h=1}^{H_g} \mathbb{1}_{(z^{(gh)}=z)} \int_0^{\tau^{(gh)}} \mathbb{1}_{(t_{i-1},t_i]} \mathbb{1}_{(X^{(gh)}(t-)=j)} \, \mathrm{d}t. \tag{6.2.6}$$

Thus imposing the likelihood principle, we might as well consider a Poisson regression consisting of independent observations

$$(O_{jk}(i|g, z))_{i=1,\ldots,K; g=1,\ldots,G; z \in \mathcal{Z}}$$

with distributions Poisson$(E_j(i|g, z)\xi_{jk}^{(g)} \mu_{jk}(t_i; \beta|z))$ under $\mathbb{P}_{(\xi,\beta)}$, where the exposure $E_j(i|g, z)$ is considered fixed. Considering the parameters β as known, the maximum likelihood estimator $\hat{\xi}^{(g)}_{jk}$ of $\xi^{(g)}_{jk}$ would be the following groupwise ratio:

$$\hat{\xi}^{(g)}_{jk} = \frac{\sum_{z\in\mathcal{Z}}\sum_{i=1}^{K} O_{jk}(i|g,z)}{\sum_{z\in\mathcal{Z}}\sum_{i=1}^{K} E_j(i|g,z)\mu(t_i;\beta|z)}.$$

For small groups, this estimator is relatively volatile implying a sub-optimal bias-variance trade-off and unsatisfying predictive performance. For successful experience rating in the classic Markov chain life insurance setting, we therefore need to perform shrinkage estimation of the group effects $\xi^{(g)}_{jk}$. This is the focal point of the next section.

6.3 A multivariate frailty and empirical Bayes approach to experience rating in the classic Markov chain life insurance setting

In this section, we extend the classic Markov chain life insurance setting to allow for (latent) group heterogeneity, hereby generalizing the novel approach of Norberg (1989b, Section 2). The resulting framework is that of multivariate frailty modeling known from survival and event history analysis – for an overview see e.g. Vaupel, Manton, and Stallard (1979), Andersen et al. (1988), and Hougaard (2000). The purpose of the extension is to obtain interpretable shrinkage estimation of the group effects by applying empirical Bayes methods.

In Subsection 6.3.1, we carry out the extension for the survival model including a survey of parts of Norberg (1989b). An interesting comparison study in relation to this is Haastrup (2000). In Subsection 6.3.2, we carry out the extension for general multi-state models. We provide a characterization of what we call the *three points of transition entanglement*, i.e. we characterize models which yield a simpler (product) structure for the relevant likelihoods. The split-up observed in Example 6.2.1 of the previous section can be considered an especially simple consequence of our characterization. Van Der Gaag et al. (2015) discuss the relation between model assumptions and estimation procedures for general multi-state models with frailty emphasizing the correlation structure between latent variables. We instead focus on the parametric structure.

A particularly simple shrinkage estimation procedure is obtained by assuming the latent variables within groups to be mutually independent and Gamma-distributed and considering the Bayes estimator under quadratic loss. We study this approach in details. While it is not original for frailty models in general, we believe that our specific presentation in a multi-state life insurance context provides new understanding concerning parametric modeling and estimation of group effects for the classic Markov chain life insurance setting.

6.3.1 Survival model

We extend the survival model to allow for latent group effects. The state space S consists of two states, $S = \{0, 1\}$, with 1 an absorbing state denoting death, and the model is fully specified through the distribution of the counting processes $N^{(gh)}$, $g = 1, \dots, G$ and $h = 1, \dots, H_g$, where $N^{(gh)}$ counts the number of deaths of insured h in group g. Details regarding the canonical construction related to the following specification can be found in Andersen et al. (1988, Chapter IX).

Let $\Theta = (\Theta^{(1)}, \dots, \Theta^{(G)})$ be a vector of independent and identically distributed strictly positive random variables. In addition to the regression parameters β, consider hyper-parameters ψ solely describing the marginal distribution of $\Theta^{(g)}$ with values in some parameter set Ψ. To be precise, we require that under the probability measure $\mathbb{P}_{\beta,\psi}$, $\Theta^{(g)}$ has distribution Π depending on ψ and not β.

To obtain a full specification (under each probability measure $\mathbb{P}_{\beta,\psi}$), we assume that conditionally on Θ, the counting processes are independent with conditional intensity processes given by

$$\lambda^{(gh)}(t; \beta|\theta) = \mathbb{1}_{(N^{(gh)}(t-)=0)}\theta^{(g)}\mu(t; \beta|z^{(gh)})$$

not depending on ψ. We see that the fixed effect $\xi_{01}^{(g)}$ has been replaced by a random effect $\theta^{(g)}$. This introduces dependency within groups and heterogeneity between groups.

Because we only observe the counting processes N and not the latent variables Θ, the marginal likelihood is of primary interest. By integrating out Θ and using the assumed (conditional) independence, the marginal likelihood reads

$$\begin{aligned}\mathcal{L}(\beta, \psi) = \prod_{g=1}^{G} \int_0^\infty \Bigg(\prod_{h=1}^{H_g} \exp\bigg\{ \int_{(0,\tau^{(gh)}]} \log\big(\theta^{(g)}\mu(t; \beta|z^{(gh)})\big) \, \mathrm{d}N^{(gh)}(t) \\ - \theta^{(g)} \int_0^{\tau^{(gh)}} \mathbb{1}_{(N^{(gh)}(t-)=0)}\mu(t; \beta|z^{(gh)}) \, \mathrm{d}t \bigg\} \Bigg) \Pi(\mathrm{d}\theta^{(g)}; \psi).\end{aligned} \tag{6.3.1}$$

To obtain maximum likelihood estimates of the regression parameters β and hyper-parameters ψ, one must maximize this expression numerically.

In Norberg (1989b), Norberg studies estimation and valuation in the above setting under the assumption of Gamma-distributed latent variables Θ, which is a family of so-called conjugate priors. In the remainder of this subsection let $\Psi = (0, \infty)$ and assume that $\Theta^{(g)} \sim \Gamma(\psi^{-1}, \psi^{-1})$. In general, we allow for scaling of the base transition intensities. Hence the restriction to Gamma distributions with mean one is necessary to avoid overparametrization. We now give a short summary of relevant aspects of Norberg (1989b) for our setting adding additional insights regarding shrinkage estimation and experience rating in practice.

Norberg shows that the present value of a lump sum of one upon death with risk period Δ in group g is approximated (as $\Delta \to 0$) by $\Delta\tilde{\Theta}^{(g)}\mu(t;\beta|z)$ where $\tilde{\Theta}$ is the Bayes estimator of Θ under quadratic loss. The following expression for $\tilde{\Theta}$ is proven in Norberg (1989b):

Proposition 6.3.1. *Assume that* $\Psi = (0,\infty)$ *and* $\Theta^{(g)} \sim \Gamma(\psi^{-1},\psi^{-1})$. *Then the Bayes estimator of* $\tilde{\Theta}$ *of* Θ *under quadratic loss is given by*

$$\tilde{\Theta}^{(g)} = \zeta^{(g)} \cdot \hat{\Theta}^{(g)} + (1-\zeta^{(g)}) \cdot 1 \tag{6.3.2}$$

where

$$\hat{\Theta}^{(g)} = \frac{\sum_{h=1}^{H_g} N^{(gh)}(\tau^{(gh)})}{\sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(N^{(gh)}(t-)=0)}\mu(t;\beta|z^{(gh)})\,\mathrm{d}t}$$

is the 'conditional' maximum likelihood estimator of the group effect for fixed regression parameters β, 1 *is exactly the prior/unconditional mean of* $\Theta^{(g)}$, *and*

$$\zeta^{(g)} = \frac{\sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(N^{(gh)}(t-)=0)}\mu(t;\beta|z^{(gh)})\,\mathrm{d}t}{\sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(N^{(gh)}(t-)=0)}\mu(t;\beta|z^{(gh)})\,\mathrm{d}t + \psi^{-1}}.$$

We interpret (6.3.2) as a (linear) credibility formula. The Bayes estimator $\tilde{\Theta}^{(g)}$ is a weighted average between the 'population average' 1 and the 'group average' $\hat{\Theta}^{(g)}$, where the credibility weight $\zeta^{(g)}$ is an increasing function of the time under risk of transition and the prior/unconditional variance ψ, respectively.

The credibility formula (6.3.2) is very appealing and can serve as a motivation for a simple shrinkage estimator of the group effect. We propose valuation for group g using the 'plug-in' transition intensities

$$t \mapsto \tilde{\Theta}^{(g)}\mu(t;\beta|z)$$

with the hyper-parameter ψ and the regression parameters β estimated from the marginal likelihood $\mathcal{L}(\beta,\psi)$. For lump sum payments with infinitesimal risk periods, it follows from the results of Norberg that this leads to predictions that are optimal in a Bayesian sense w.r.t. quadratic loss. For other types of payments, e.g. life annuities, this is not the case. It is in general difficult to quantify the resulting prediction errors. Thus a careful analysis of the predictive performance of this simple shrinkage estimator is required. We return to this discussion at the end of Subsection 6.3.2.

Norberg does not discuss estimation of the regression parameters when only summary statistics such as the number of deaths and exposures are available. To our belief, the following unsurprising link to multivariate negative binomial regressions is therefore new.

Let $0 = t_0 < t_1 < \cdots < t_K = \max_{g,h} \tau^{(gh)}$ be a grid, and assume that the base transition intensity μ is continuous from the left and piecewise constant on this grid. In the same manner as in Subsection 6.2.2 and Subsection 6.2.3, the marginal likelihood (6.3.1) simplifies and becomes proportional to the likelihood of independent multivariate negative binomial random variables.

To be precise, let $O(i|g,z)$ be the number of deaths for time interval i for insured in group g with covariate z, and let $E(i|g,z)$ be the total time spent alive (exposure) in time interval i for insured in group g with covariate z, confer with (6.2.5)–(6.2.6). Impose the alternative assumption that conditionally on Θ, the number of deaths are independent with Poisson($\Theta^{(g)}E(i|g,z)\mu(t_i;\beta|z)$)-distributions, where the exposures are considered fixed. Consequently, the marginal likelihood of the number of deaths is proportional to (6.3.1). When only summary statistics such as the number of deaths and exposures are available, likelihood estimation of the hyper-parameter ψ and the regression parameters β can therefore be based on Poisson-Gamma mixture (i.e. multivariate negative binomial) regressions.

To our knowledge, maximum likelihood estimation for multivariate negative binomial regressions has not currently been fully implemented in any accessible R-package. In a later numerical example for disability insurance, see Section 6.4, we apply a simple EM-algorithm proposed by Ghitany et al. (2012).

6.3.2 Multivariate extension

We extend the classic Markov chain life insurance setting to allow for latent group effects. We mimic the approach of the previous subsection adapting it to the multi-state setting. Consider the finite state space $S = \{0, 1, \ldots, J\}$ and jump processes $X^{(gh)}$, $g = 1, \ldots, G$ and $h = 1, \ldots, H_g$, with corresponding multivariate counting processes $N^{(gh)}$. For each group g of insured we now have p strictly positive (latent) random variables $\Theta^{(g)} = (\Theta_1^{(g)}, \ldots, \Theta_p^{(g)})$. We assume that latent variables between groups are mutually independent and identically distributed, while initially, no such assumptions are made for latent variables within groups. In particular, latent variables within a group can be dependent. The prior distribution of $\Theta^{(g)}$ is denoted Π and depends on the hyper-parameters ψ and not the regression parameters β.

Set $\mathcal{S} = \{(j,k) \in S^2 : k \neq j\}$. This is the set of possible transitions. The set of transitions affected by the latent variables is denoted $\mathcal{J} \subset \mathcal{S}$. The latent variables are not allowed to affect the same transitions, requiring $p \leq \#\mathcal{J}$. To model which transitions are affected by which latent variable, we introduce a surjection $\pi : \mathcal{J} \to \{1, \ldots, p\}$. If $\pi(0,1) = i$, this means $\Theta_i^{(g)}$ affects the transition $(0,1)$, while the preimage of i under π, i.e. $\pi^{-1}(i) \subset \mathcal{J}$, are all the transitions affected by the i'th latent variable $\Theta_i^{(g)}$.

To obtain a full specification (under each probability measure $\mathbb{P}_{\beta,\psi}$), we assume that conditionally on Θ, the multivariate counting processes are independent with conditional intensity processes given by

$$\lambda_{jk}^{(gh)}(t;\beta|\theta) = \begin{cases} \mathbb{1}_{(X^{(gh)}(t-)=j)}\mu_{jk}(t;\beta|g,z^{(gh)}) & (j,k)\in\mathcal{J}^{\mathsf{c}} \\ \mathbb{1}_{(X^{(gh)}(t-)=j)}\theta_{\pi(j,k)}^{(g)}\mu_{jk}(t;\beta|z^{(gh)}) & (j,k)\in\mathcal{J} \end{cases}$$

not depending on ψ. If we compare to Subsection 6.2.3, we see that for the transitions $\mathcal{J}$, the fixed group effects have been replaced by random group effects with prior Π. For the transitions $\mathcal{J}^{\mathsf{c}}$ we have not specified the group effects, except that they are assumed to be non-random. While the groups remain independent, the jump processes $X^{(g1)},\ldots,X^{(gH_g)}$ are now only independent and Markovian conditionally on $\Theta^{(g)}$. This introduces dependency within groups and heterogeneity between groups. If $S=\{0,1\}$, $\mathcal{J}=\{(0,1)\}$, and $\mu_{10}=0$, we recover the extended survival model of Subsection 6.3.1. If $\mathcal{J}=\emptyset$, such that no transitions are affected by latent effects, we recover the classic Markov chain life insurance setting. In the remainder of the section, we assume $\mathcal{J}\neq\emptyset$.

Disentanglement

We proceed with an investigation of the (product) structure of the marginal likelihood. At the end of Subsection 6.2.1, see in particular Example 6.2.1, we showed that distinct parametrization of the transition intensities affects sufficiency of statistics and simplifies maximum likelihood estimation for the classic Markov chain life insurance setting by splitting the original inferential problem into a number of sub-problems. This is important to the actuarial practitioner, as it allows him/her to take a marginal approach to data collection and risk estimation. In this subsection, we adapt the characterization for the classic setting to the extended setting.

As the groups are independent, the number of groups is not important for the (product) structure of the likelihoods. Thus in the following, we take $G=1$ and suppress the group number g in the notation.

By integrating out the latent variables Θ, the marginal likelihood reads

$$\mathcal{L}(\beta,\psi) = \prod_{(j,k)\in\mathcal{J}^{\mathsf{c}}}\mathcal{L}_{jk}(\beta)\int_{(0,\infty)^p}\Bigg(\prod_{(j,k)\in\mathcal{J}}\mathcal{L}_{jk}(\beta|\theta)\Bigg)\Pi(\mathrm{d}\theta;\psi). \tag{6.3.3}$$

Here for $(j,k)\in\mathcal{J}$,

$$\mathcal{L}_{jk}(\beta|\theta) = \prod_{h=1}^{H}\exp\Bigg\{\int_{(0,\tau^{(h)}]}\log\big(\theta_{\pi(j,k)}\mu_{jk}(t;\beta|z^{(h)})\big)\,\mathrm{d}N_{jk}^{(h)}(t) \\ -\theta_{\pi(j,k)}\int_0^{\tau^{(h)}}\mathbb{1}_{(X^{(h)}(t-)=j)}\mu_{jk}(t;\beta|z^{(h)})\,\mathrm{d}t\Bigg\},$$

while for $(j,k) \in \mathcal{J}^{\mathrm{c}}$,

$$\mathcal{L}_{jk}(\beta) = \prod_{h=1}^{H} \exp\left\{ \int_{(0,\tau^{(h)}]} \log\left(\mu_{jk}(t;\beta|z^{(h)})\right) \mathrm{d}N_{jk}^{(h)}(t) \right.$$
$$\left. - \int_0^{\tau^{(h)}} \mathbb{1}_{(X^{(h)}(t-)=j)} \mu_{jk}(t;\beta|z^{(h)}) \,\mathrm{d}t \right\}.$$

For the transitions affected by the latent variables, i.e. the transitions $\mathcal{J}$, the original product structure is potentially ruined. By inspection of (6.3.3), the (product) structure of this likelihood is determined by the following characteristics:

1. The surjection π determining which transitions have the same random effect, which we denote as the *first point of transition entanglement.*
2. The common (distinct) parametrization of the base transition intensities μ w.r.t. the regression parameters β, confer with the classic setting. This we denote as the *second point of transition entanglement.*
3. The prior distribution through:
   - Dependency between latent variables given by the (product) structure of the prior distribution Π.
   - The common (distinct) parametrization of the prior distribution $\Pi(\cdot\,;\psi)$ w.r.t. the hyper-parameters ψ.

Typically, dependency between latent variables and common parametrization of the prior distributions appear together (common parameters describing the correlations), so we combine the 3rd and 4th characteristic into a single characteristic: the *third point of transition entanglement.* Note that the first point of transition entanglement is essentially a special case of the third point because full dependency induce the same structure as the surjection π. On the other hand, it is often useful from a practical modeling perspective to be able to distinguish between full dependency and all kinds of other dependency. This is also the case here.

To characterize how the first, second, and third point of transition entanglement affect the (product) structure of the marginal likelihood, we first have to describe the parametric structures.

Let $B = (B_i)_{i=0,\ldots,p'}$ be the finest partition of $\mathcal{S}$, with B_0 possibly empty, such that the regression parameter set $\mathcal{B}$ can be written as

$$\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 \times \cdots \times \mathcal{B}_{p'},$$

and such that for any $(j,k) \in B_0$, the base transition intensity μ_{jk} does not depend on $\beta \in \mathcal{B}$, and for any $(j,k) \in B_i$, the base transition intensity μ_{jk} only depends on $\beta \in \mathcal{B}$ through $\beta_i \in \mathcal{B}_i$, i.e. for all $\beta, \beta' \in \mathcal{B}$ with $\beta_i = \beta'_i$ it holds that $\mu_{jk}(\cdot\,;\beta) = \mu_{jk}(\cdot\,;\beta')$. The partition B allows us to distinguish between transitions for which the base transition intensities are distinctly parametrized. To emphasize this, we sometimes write $\mu_{jk}(\cdot)$ for $(j,k) \in B_0$ and $\mu_{jk}(\cdot\,;\beta_i)$ for $(j,k) \in B_i$.

Let $L = (L_d)_{d=0,\ldots,p''}$ be the finest partition of $\{1,\ldots,p\}$, with L_0 possible emtpy, such that the hyper-parameter set Ψ can be written as

$$\Psi = \Psi_1 \times \Psi_2 \times \cdots \times \Psi_{p''}$$

and the prior distribution Π takes the product form

$$\Pi(\cdot\,;\psi) = \Pi_0 \otimes \Pi_1 \otimes \cdots \otimes \Pi_{p''}$$

where Π_d is the (simultaneous) prior distribution of $(\Theta_i)_{i \in L_d}$ only depending on the hyper-parameters ψ through ψ_d (with Π_0 not depending on ψ). We see that the partition L decomposes the latent variables into independent vectors (due to the product form of the prior distribution) with distinctly parametrized prior distributions. To emphasize this, we sometimes write $\Pi_0(\cdot)$ and $\Pi_i(\cdot\,;\psi_i)$.

We may combine the partition L of $\{1,\ldots,p\}$ with the surjection π to form a partition $A = (A_i)_{i=0,\ldots,p''}$, with A_0 possibly empty, of $\mathcal{J}$ (the set of transitions affected by latent variables) by setting $A_i = \pi^{-1}(L_i)$. The partition A allows us to distinguish between transitions for which the relevant latent variables (random effects) are independent with distinctly parametrized prior distributions.

Bundling together $(B_i)_{i=1,\ldots,p'}$ and $(A_i)_{i=1,\ldots,p''}$, we construct a partition $D = (D_d)_{d=1,\ldots,q}$ of the set of possible transitions $\mathcal{S}$ as the finest partition which is coarser than both A and B. This is the unique finest partition which satisfies that if $(j,k) \in D_d$ and $(j',k') \in D_{d'}$ for $d \neq d'$, then

- The base transitions μ_{jk} and $\mu_{j'k'}$ are distinctly parametrized w.r.t. the regression parameters β.
- Whenever $(j,k) \in \mathcal{J}$ and $(j',k') \in \mathcal{J}$, the latent variables (random effects) affecting the transitions, i.e. $\Theta_{\pi(j,k)}$ and $\Theta_{\pi(j',k')}$, are independent with distinctly parametrized prior distributions w.r.t. the hyper-parameters ψ.

Thus D allows us to *distinguish* between transitions taking all three points of transition entanglement into account. This leads to the following notion:

Definition 6.3.2. *We call the partition D the distinguisher of $\mathcal{S}$ (based on B and A).*

Before we can state the main result, we need to develop a final concept. It will prove useful to introduce the mappings which identify a certain vector of hyper-parameters or regression parameters with that element of the partition D which in particular contains the transitions affected (directly, or indirectly through the random effects) by this vector of parameters. These mappings, which we denote *identifiers*, are defined as follows:

Definition 6.3.3. *The unique mapping $I_{\mathcal{B}} : \{1, \ldots, p'\} \to D$ where*

$$I_{\mathcal{B}}(i) = D_d \quad \textit{if and only if} \quad B_i \subset D_d$$

is denoted the identifier of $\mathcal{B}$ (from D).

Definition 6.3.4. *The unique mapping $I_{\Psi} : \{1, \ldots, p''\} \to D$ where*

$$I_{\Psi}(i) = D_d \quad \textit{if and only if} \quad A_i \subset D_d$$

is denoted the identifier of Ψ (from D).

Because the distinguisher in particular partitions the set of possible transitions such that latent variables affecting transitions in different elements of the partition are independent, we can rewrite the marginal likelihood as

$$\mathcal{L}(\beta, \psi) = \prod_{d=1}^{q} \mathcal{L}_d(\beta, \psi)$$

with $\mathcal{L}_d$ the likelihood involving the transitions D_d given by

$$\mathcal{L}_d(\beta, \psi) = \prod_{(j,k) \in D_d \cap \mathcal{J}^c} \mathcal{L}_{jk}(\beta) \int \Bigg(\prod_{(j,k) \in D_d \cap \mathcal{J}} \mathcal{L}_{jk}(\beta|\theta) \Bigg) \tilde{\Pi}_d(\mathrm{d}\theta; \psi)$$

where $\tilde{\Pi}_d = \bigotimes_{i \in I_{\Psi}^{-1}(D_d)} \Pi_i$ is the (simultaneous) prior distribution of the latent variables (random effects) affecting the transitions D_d. Using this representation, we are now ready to state the main result which yields the desired characterization of the marginal likelihood as a product of distinctly parametrized terms:

Theorem 6.3.5. *The d'th term $\mathcal{L}_d(\beta, \psi)$ of the marginal likelihood*

$$\mathcal{L}(\beta, \psi) = \prod_{d=1}^{q} \mathcal{L}_d(\beta, \psi)$$

only depends on (β, ψ) through $(\{\beta_i : I_{\mathcal{B}}(i) = D_d\}, \{\psi_i : I_{\Psi}(i) = D_d\})$. In particular, the q terms of the likelihood are distinctly parametrized w.r.t. the hyper-parameters ψ as well as the regression parameters β.

Proof. The result is an immediate consequence of the precise construction of the distinguisher D and its resulting properties. $\square$

It follows from the theorem that the original inferential problem splits into q simpler sub-problems in the twofold sense that

1. Maximum likelihood estimation of $(\{\beta_i : I_{\mathcal{B}}(i) = D_d\}, \{\psi_i : I_\Psi(i) = D_d\})$ can be based on the simpler term $\mathcal{L}_d$.

2. Sufficient statistics for $(\{\beta_i : I_{\mathcal{B}}(i) = D_d\}, \{\psi_i : I_\Psi(i) = D_d\})$ are given by

$$\left(\mathbb{1}_{(X^{(h)}(t)=j)}, N_{jk}^{(h)}(t)\right)_{h=1,\ldots,H;t\in[0,\tau^{(h)}]}$$

for $(j,k) \in D_d$ rather than $(j,k) \in \mathcal{S}$.

This has important consequences for and connotations to the workflow of the actuarial practitioner. The simplicity, and thus also the interpretability, of a class of models is seen to be partly associated to the fineness/coarseness of the distinguisher D. The simple shrinkage estimators, which we derive in the next subsection, serve as a good example of how studying the likelihood structure through the distinguisher motivates an estimation procedure.

The following example illustrates the concepts of distinguishers and identifiers and the consequences of Theorem 6.3.5.

Example 6.3.6. Let $S = \{0, 1, 2\}$ with 0 denoting active, 1 denoting disabled, and 2 denoting dead. Let $p = 2$ (two latent variables $\Theta = (\Theta_1, \Theta_2)$), such that conditionally on Θ, the jump processes are independent and Markovian with transition intensities as in Figure 6.2. In comparison to the classic Markovian disability model of Example 6.2.1, dependency within groups and heterogeneity between groups has been introduced w.r.t. disability and mortality as disabled, which are the main biometric risk factors for disability insurance. As in Example 6.2.1, we allow the mortalities to be commonly parametrized by $\beta_2 \in \mathcal{B}_2$, which is e.g. the case if the mortality as disabled is estimated relatively to the mortality as active, while the disability rate is distinctly parametrized by $\beta_1 \in \mathcal{B}_1$ (with $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$). Consequently,

$$B_0 = \{(2,0), (2,1), (1,0)\}, \qquad B_1 = \{(0,1)\}, \qquad B_2 = \{(0,2), (1,2)\}.$$

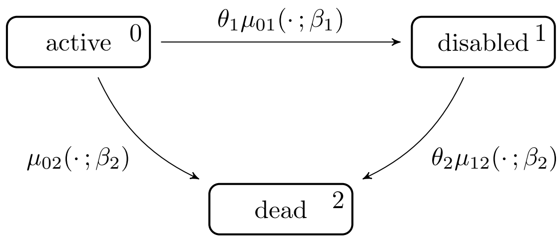

Figure 6.2: *Extended disability model without recovery, where conditionally on $\Theta = \theta$, the disability rate is $\theta_1\mu_{01}(\cdot\,;\beta_1)$, the mortality as disabled is $\theta_2\mu_{12}(\cdot\,;\beta_2)$, and the mortality as active is simply $\mu_{02}(\cdot\,;\beta_2)$.*

If no assumption is made regarding the dependency between latent variables and the (distinct) parametrization of their prior distributions, then

$$A_0 = \emptyset, \qquad A_1 = \mathcal{J} = \{(0,1),(1,2)\}.$$

The distinguisher D (based on B and A) is therefore given by

$$D_1 = \{(0,1),(0,2),(1,2)\}, \qquad D_2 = \{(1,0)\}, \qquad D_3 = \{(2,0)\}, \qquad D_4 = \{(2,1)\}.$$

see also Figure 6.3. The identifiers are given by

$$I_{\mathcal{B}}(1) = I_{\mathcal{B}}(2) = I_{\Psi}(1) = D_1.$$

It follows from Theorem 6.3.5 that all parameters (β,ψ) must be estimated simultaneously involving data related to the transitions D_1.

If we compare this to Example 6.2.1, we see that in both cases the common parametrization of the mortalities 'entangles' the transitions $(0,2)$ and $(1,2)$. But the inclusion of (dependent and/or commonly parametrized) random effects also 'entangles' the transitions $(0,1)$ and $(1,2)$. Therefore, simultaneous estimation of all parameters involving the transitions $(0,1)$, $(0,2)$, and $(1,2)$ is now necessary. Note that even if Θ_1 and Θ_2 are assumed independent, distinct parametrization of their prior distributions is still needed for disentanglement. ∘

Shrinkage estimation

Previously we determined how the complexity of the estimation of the hyperparameters and regression parameters relates to the dependency between the latent variables and the parametric structure of the base transition intensities and the prior distributions. We now provide a series of sufficient conditions which guarantee significant simplifications and motivates a simple shrinkage estimation procedure based on empirical Bayes methods.

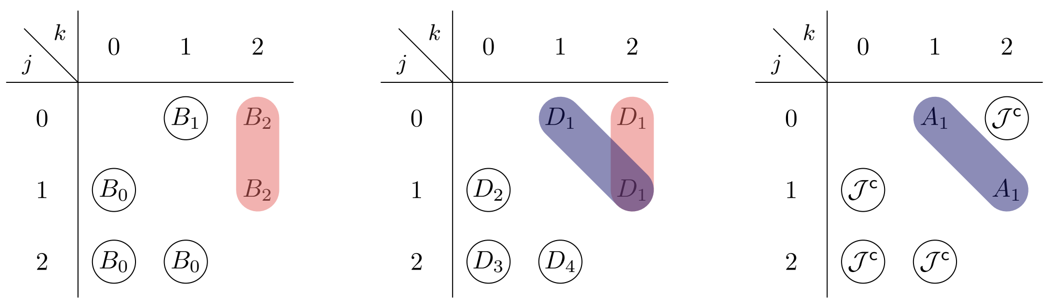

Figure 6.3: *Construction of the distinguisher D (based on B and A). It appears by combining the two overlays representing the second and the combined first and third points of transition entanglement.*

Empirical Bayes Empirical Bayes estimation involves Bayes estimation of the random effects. Taking the hyper-parameters and regression parameters for given, the following extension of Proposition 6.3.1 provides sufficient conditions to obtain explicit shrinkage estimators of the group effects:

Proposition 6.3.7. *Assume that the latent variables within groups are independent and that $\Theta_i^{(g)}$ is Gamma-distributed with mean 1 and variance $v_i(\psi)$. Then for the Bayes estimator $\tilde{\Theta}$ of Θ under quadratic loss it holds that*

$$\tilde{\Theta}_i^{(g)} = \zeta_i^{(g)} \cdot \hat{\Theta}_i^{(g)} + (1 - \zeta_i^{(g)}) \cdot 1. \tag{6.3.4}$$

Here

$$\hat{\Theta}_i^{(g)} = \frac{\sum_{(j,k)\in\pi^{-1}(i)} \sum_{h=1}^{H_g} N_{jk}^{(gh)}(\tau^{(gh)})}{\sum_{(j,k)\in\pi^{-1}(i)} \sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(X^{(gh)}(t-)=j)} \mu_{jk}(t;\beta|z^{(gh)})\,\mathrm{d}t}$$

is the 'conditional' maximum likelihood estimator of the group effect for transitions $\pi^{-1}(i)$ for fixed regression parameters β, 1 is exactly the prior/unconditional mean of $\Theta_i^{(g)}$, and

$$\zeta_i^{(g)} = \frac{\sum_{(j,k)\in\pi^{-1}(i)} \sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(X^{(gh)}(t-)=j)} \mu_{jk}(t;\beta|z^{(gh)})\,\mathrm{d}t}{\sum_{(j,k)\in\pi^{-1}(i)} \sum_{h=1}^{H_g} \int_0^{\tau^{(gh)}} \mathbb{1}_{(X^{(gh)}(t-)=j)} \mu_{jk}(t;\beta|z^{(gh)})\,\mathrm{d}t + \frac{1}{v_i(\psi)}}.$$

Proof. Due to the assumption of independent latent variables, the result follows immediately by an application of the proof technique from the extended survival model, see Norberg (1989b, equations (2.7)-(2.9)). □

The requirement that $\Theta_i^{(g)}$ has mean one is imposed to avoid overparametrization, see also the discussion in Subsection 6.3.1.

Note that to calculate $\tilde{\Theta}_i^{(g)}$, only the data

$$\left(\mathbb{1}_{(X^{(gh)}(t)=j)}, N_{jk}^{(gh)}(t)\right)_{h=1,\dots,H_g;t\in[0,\tau^{(gh)}]}$$

for $(j,k) \in \pi^{-1}(i)$ rather than $(j,k) \in \mathcal{S}$ is needed.

The interpretation of (6.3.4) is the same as in the extended survival model. It is a (linear) credibility formula, so that $\tilde{\Theta}_i^{(g)}$ is a weighted average between the 'population' average 1 and the 'group average' $\hat{\Theta}_i^{(g)}$, where the credibility weight $\zeta_i^{(g)}$ is an increasing function of the time under risk of transition and the prior/unconditional variance $v_i(\psi)$, respectively.

We now discuss estimation of the hyper-parameters and regression parameters. In the empirical Bayes approach, these parameters are estimated using the marginal likelihood. From Example 6.3.6, we know that assuming the latent variables within groups to be independent is not sufficient to avoid the necessity of a simultaneous estimation of all parameters. The following corollary to Theorem 6.3.5 provides sufficient conditions on the parametric structures to guarantee that the original inferential problem splits into sub-problems at most involving one of the random effects.

Corollary 6.3.8. *Assume that*

1. *The latent variables within groups are independent with prior marginal distributions* Π_i, $i = 1, \dots, p$.
2. *The prior marginal distributions are distinctly parametrized, so that* Π_i *only depends on* $\psi \in \Psi$ *through* $\psi_i \in \Psi_i$ $(\Psi = \Psi_1 \times \Psi_2 \times \cdots \times \Psi_p)$.
3. *If transitions* $(j,k), (j',k') \in \mathcal{S}$ *are affected by distinct latent variables or one of the transitions is unaffected by the latent variables, the corresponding base transition intensities are distinctly parametrized. (In other words, B is finer than A.)*

Write β_i *for the regression parameters parametrizing the base transition intensities for transitions affected by the i'th latent variable, i.e. for transitions* $(j,k) \in \pi^{-1}(i)$, *and write* β_0 *for the regression parameters parametrizing the base transition intensities for transitions unaffected the latent variables, i.e. for transitions* $(j,k) \in \mathcal{J}^{\mathsf{c}}$. *Then maximum likelihood estimation of* (β_i, ψ_i) *only requires maximization of the simpler term*

$$\mathcal{L}_i(\beta_i, \psi_i) = \prod_{g=1}^{G} \int_0^\infty \Bigg(\prod_{(j,k)\in\pi^{-1}(i)} \mathcal{L}_{jk}^{(g)}(\beta_i | \theta_i) \Bigg) \Pi_i(\mathrm{d}\theta_i; \psi_i),$$

while maximum likelihood estimation of β_0 *only requires maximization of the simpler term*

$$\mathcal{L}_0(\beta_0) = \prod_{g=1}^{G} \prod_{(j,k)\in\mathcal{J}^{\mathsf{c}}} \mathcal{L}_{jk}^{(g)}(\beta_0).$$

Here

$$\begin{aligned}
\mathcal{L}_{jk}^{(g)}(\beta_i|\theta_i) = \prod_{h=1}^{H_g} \exp\Bigg\{ &\int_{(0,\tau^{(gh)}]} \log\big(\theta_i \mu_{jk}(t;\beta_i|z^{(gh)})\big)\, \mathrm{d}N_{jk}^{(gh)}(t) \\
&- \theta_i \int_0^{\tau^{(gh)}} \mathbb{1}_{(X^{(gh)}(t-)=j)} \mu_{jk}(t;\beta_i|z^{(gh)})\, \mathrm{d}t \Bigg\}, \\
\mathcal{L}_{jk}^{(g)}(\beta_0) = \prod_{h=1}^{H_g} \exp\Bigg\{ &\int_{(0,\tau^{(gh)}]} \log\big(\mu_{jk}(t;\beta_0|g, z^{(gh)})\big)\, \mathrm{d}N_{jk}^{(gh)}(t) \\
&- \int_0^{\tau^{(gh)}} \mathbb{1}_{(X^{(gh)}(t-)=j)} \mu_{jk}(t;\beta_0|g, z^{(gh)})\, \mathrm{d}t \Bigg\}.
\end{aligned}$$

Furthermore,

$$\left(\mathbb{1}_{(X^{(gh)}(t)=j)}, N_{jk}^{(gh)}(t)\right)_{g=1,\dots,G;h=1,\dots,H_g;t\in[0,\tau^{(gh)}]}$$

for $(j,k) \in \pi^{-1}(i)$ *are sufficient statistics for* (β_i, ψ_i)*, and*

$$\left(\mathbb{1}_{(X^{(gh)}(t)=j)}, N_{jk}^{(gh)}(t)\right)_{g=1,\dots,G;h=1,\dots,H_g;t\in[0,\tau^{(gh)}]}$$

for $(j,k) \in \mathcal{J}^{\mathrm{c}}$ *are sufficient statistics for* β_0*.*

Simple shrinkage Under the restrictive assumptions of Proposition 6.3.7 and Corollary 6.3.8, with $\Psi_i = (0,\infty)$ and $v_i(\psi) = \psi_i$, we obtain the following empirical Bayes estimation procedure:

- Estimate β_0 by maximizing $\mathcal{L}_0$. This yields maximum likelihood estimates $\hat{\beta}_0$.
- For each $i = 1,\dots,p$ using the data

  $$\left(\mathbb{1}_{(X^{(gh)}(t)=j)}, N_{jk}^{(gh)}(t)\right)_{g=1,\dots,G;h=1,\dots,H_g;t\in[0,\tau^{(gh)}]}, \qquad (j,k) \in \pi^{-1}(i),$$

  - Estimate (β_i, ψ_i) by maximizing $\mathcal{L}_i$. This yields maximum likelihood estimates $(\hat{\beta}_i, \hat{\psi}_i)$.
  - For $g = 1,\dots,G$ estimate $\Theta_i^{(g)}$ by plugging $(\hat{\beta}_i, \hat{\psi}_i)$ into (6.3.4). This yields shrinkage estimates $\tilde{\Theta}_i^{(g)}$ of the group effects.

A simple shrinkage estimation procedure is obtained by assuming the base transition intensities to be continuous from the left and piecewise constant on some grid $0 = t_0 < t_1 < \cdots < t_K = \max_{g,h} \tau^{(gh)}$. Let $O_{jk}(i|g,z)$ be the number of transitions from j to k in the time-interval i for insured in group g with covariate z, and let $E_j(i|g,z)$ be the total time spent (exposure) in state j in time interval i for insured in group g with covariate z, confer with (6.2.5)–(6.2.6). It follows that the occurrences and exposures

$$(E_j(i|g,z), O_{jk}(i|g,z))_{i=1,\dots,K;g=1,\dots,G;z\in\mathcal{Z}}$$

for the transitions $(j,k) \in \pi^{-1}(i)$ are sufficient to calculate $\tilde{\Theta}_i$ and sufficient statistics for (β_i, ψ_i). Furthermore, by similar arguments as in Subsection 6.3.1, the marginal likelihood $\mathcal{L}_i$ is proportional to that of a Poisson-Gamma mixture (i.e. multivariate negative binomial) regression, where conditionally on Θ_i, all the occurrences O_{jk} are independent with Poisson$(\Theta_i^{(g)} E_j(i|g,z)\mu_{jk}(t_i;\beta|z))$-distributions, where the exposures are considered fixed and $(j,k) \in \pi^{-1}(i)$. Hence estimation of (β_i, ψ_i) can proceed in similar fashion as in the extended survival model. For transitions $(j,k) \in \mathcal{J}^{\mathrm{c}}$ unaffected by the latent variables, the regression parameters β_0 can be estimated using Poisson regressions as in the classic setting.

Prediction We propose valuation of transition payments and sojourn payments for group g in an 'as if' Markov setting using the 'plug-in' transition intensities

$$t \mapsto \tilde{\Theta}^{(g)}_{\pi(j,k)} \mu_{jk}(t; \beta | z)$$

for $(j,k) \in \mathcal{J}$ and

$$t \mapsto \mu_{jk}(t; \beta | g, z)$$

for $(j,k) \in \mathcal{J}^{\mathsf{c}}$ and with the hyper-parameters ψ and the regression parameters β estimated as described above. In general, the resulting prediction errors are non-traceable, so a careful analysis of the predictive performance is required. The performance measure has to be chosen with the specific application in mind. For disability insurance, the disability durations are reasonable proxies for the liabilities in a low interest rate environment, and this suggests measuring the predictive performance by comparing observed disability durations with predicted disability durations. A numerical example is given in Section 6.4.

Discussion We end this subsection with a brief discussion on the simple shrinkage estimation procedure. In general, we are not guaranteed to obtain a better bias/variance than using classic methods (such as maximum likelihood estimation without random effects). Regarding the assumptions of distinct parametrization, we stress that this only affects efficiency of the maximum likelihood estimators. If the assumption is violated, the resulting likelihoods can still be viewed as so-called partial likelihoods, see also Andersen et al. (1988, pp. 150–152), in which case the maximum likelihood estimators are consistent under suitable regularity conditions, see Wong (1986). The assumption of independent (Gamma-distributed) latent variables within groups is more critical. Consider the disability model without recovery, and let the grouping be given by employer. Certain traits leading to a simultaneous increased risk of death as well as disability (e.g. smoking) are then more prevalent in some groups than others (e.g. due to company health policies). If these traits are not included as covariates, it is reasonable to expect that estimation of the group effects allowing for positively correlated rather than independent latent variables leads to improved predictions (borrowing strength).

6.4 Numerical example

We now illustrate the applicability of the simple shrinkage estimators proposed in the last subsection through a small numerical example using simulated data. Simulation and numerics are carried out using the programming language R. The setup of interest is that of a classic Markovian disability model without recovery, see Figure 6.1, where we focus on shrinkage estimation of the disability rate and the mortality as disabled. The portfolio to be simulated is taken to consist of $G = 100$

independent groups of insured observed in $\tau = 15$ years. In the following, we use age rather than calendar time as the time scale, adjusting the previous results of Section 6.2 and Section 6.3 as necessary.

The base transition intensities are inspired by the Danish G82 tables and take the following form:

$$\begin{aligned} \mu_{02}(x) &= 0.0005 + 10^{5.88+0.038x-10}, \\ \log(\mu_{01}(x;\beta_1)) &= \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \beta_1, \\ \log(\mu_{12}(x;\beta_2)) &= \begin{bmatrix} 1 & x \end{bmatrix} \beta_2, \end{aligned}$$

for age $x \in [0, 60]$ years and regression parameters $\beta_1 \in \mathbb{R}^3$ and $\beta_2 \in \mathbb{R}^2$. The true values of β_1 and β_2 are taken to be

$$\beta_1^\mathsf{T} = \begin{bmatrix} -3.2 & -0.025 & 0.0006 \end{bmatrix}, \qquad \beta_2^\mathsf{T} = \begin{bmatrix} -7.25 & 0.07 \end{bmatrix}.$$

Note that the transition intensities are distinctly parametrized and the mortality as active is assumed to be known to the insurer. We introduce group effects $\Theta^{(g)} = (\Theta_1^{(g)}, \Theta_2^{(g)})$ with Θ_1 affecting the disability rates and Θ_2 affecting the mortalities as disabled, compare also with Example 6.3.6.

6.4.1 Simulation and data aggregation

Having partly specified the model of interest, we now turn to simulation of the remaining parameters and the dataset. All simulations are done independently. In general, we do not believe the group effects to be Gamma-distributed even though this distributional assumption is used to motivate the simple shrinkage estimation procedure. Therefore, we simulate the group effects as follows:

$$\Theta_1^{(g)} \sim \text{Uniform}[0.55, 1.45], \qquad \Theta_2^{(g)} \sim \text{Uniform}[0.75, 1.25].$$

Next, we sample the number of insured H_g in each group uniformly from the set $\{32, 36, \dots, 320\}$, and for each insured we sample his age in years at entry into the portfolio uniformly from the set $\{18, 19, \dots, 45\}$. For each insured, we then simulate the path of a Markov jump process with initial state (at entry) active and transition intensities corresponding to the true model specified above. This involves the simulation of inhomogeneous Poisson processes, which we carry out using the algorithm described in Lewis and Shedler (1979).

We randomly split the simulated dataset into two parts: a training dataset and a test dataset. The training dataset consists of three-fourths of the insured in each group, while the test dataset contains the remaining insured. In an attempt to recreate the data structure often present in actuarial practice, the training dataset is transformed into occurrences $O_{jk}(x|g)$ and exposures $E_j(x|g)$, so that e.g. $O_{01}(x|g)$

is the number of disabilities for group g in the age-interval $(x-1,x]$. The insurance contracts we have in mind are disability annuities. In a low interest rate environment and in the context of pricing and reserving of disability annuities, a reasonable proxy to assess the predictive performance is the total disability durations for the groups. Therefore, the test dataset is transformed such that it also contains total disability durations per group.

6.4.2 Estimation

Based on the previously established links to Poisson regressions and multivariate negative binomial regressions, we estimate the parameters (β^0, β) using the training dataset as follows.

For the disability rate, i.e. the group effects Θ_1 and the regression parameters β_1, we consider three different estimation procedures, starting with a 'standard' maximum likelihood estimation which does not take group effects into account. Here we set Θ_1 to one, while we estimate β_1 by maximizing the likelihood corresponding to a Poisson regression consisting of independent observations

$$(O_{01}(x))_x, \qquad O_{01}(x) = \sum_g O_{01}(x|g),$$

with canonical log-link function and linear predictor (mean function)

$$\eta(x) = \log E_0(x) + \Big[1 \quad (x-0.5) \quad (x-0.5)^2\Big]\beta_1, \qquad E_0(x) = \sum_g E_0(x|g).$$

The estimation is carried out in R using the *glm* function. The next procedure is a 'fixed effect' maximum likelihood estimation, which takes group effects into account. Here we restrict β_1 to the set $\{0\} \times \mathbb{R}^2$ to avoid overfitting and estimate Θ_1 and β_1 by maximizing the likelihood corresponding to a Poisson regression consisting of independent observations

$$(O_{01}(x|g))_{x;g}$$

with canonical log-link function and linear predictor (mean function)

$$\eta(x|g) = \log E_0(x|g) + \log \Theta_1^{(g)} + \Big[1 \quad (x-0.5) \quad (x-0.5)^2\Big]\beta_1.$$

This estimation is also carried out in R using the *glm* function.

Finally, we consider 'shrinkage' estimation as discussed at the end of the previous section. We assume that $\Theta_1^{(1)}, \dots, \Theta_1^{(100)}$ are iid $\Gamma(\psi_1^{-1}, \psi_1^{-1})$-distributed, $\psi_1 \in (0,\infty)$, and that conditionally on Θ_1, the occurrences $(O_{01}(x|g))_{x;g}$ are independent with conditional distributions

$$O_{01}(x|g) \mid \Theta_1 \sim \text{Poisson}\Big(E_0(x|g)\Theta_1^{(g)} \exp\Big\{\Big[1 \quad (x-0.5) \quad (x-0.5)^2\Big]\beta_1\Big\}\Big).$$

This corresponds to a multivariate negative binomial regression. We perform maximum likelihood estimation of (β_1, ψ_1), implementing the EM-algorithm proposed by Ghitany et al. (2012) in R. Finally, shrinkage estimates of Θ_1 are obtained by plugging the maximum likelihood estimates of (β_1, ψ_1) into the Bayes estimator under quadratic loss as outlined in the previous section.

In similar fashion, we perform 'standard', 'fixed effect', and 'shrinkage' estimation of the mortality as disabled, i.e. the group effects Θ_2 and regression parameters β_2. The resulting estimated disability rates and estimated mortalities as disabled for three selected groups are presented in Figure 6.4 together with the corresponding true rates and raw (observed) rates. For the disability rates, the fixed effect fit and shrinkage fit clearly outperform the standard fit, while the standard fit and the shrinkage fit outperform the fixed effect fit for the mortalities as disabled. This is not surprising given that the true group effects are assumed to be higher regarding disability than mortality as disabled, equally if not more important, that the disability rates are estimated based on significantly more data than the mortalities as disabled, reflecting the usual composition of a disability insurance portfolio.

6.4.3 Predictive performance

Based on the estimates from the training dataset we obtain three Markov model proposals with transition intensities given by the standard fit, fixed effect fit, and shrinkage fit, respectively. In regards to pricing, it is essential to obtain precise predictions of the disability duration (from entry until right-censoring), while it is also important in the context of reserving to obtain precise predictions of the disability duration given disability (from disability onset until right-censoring), the so-called conditional disability duration. Hence for each insured in the test dataset, we predict the disability duration as well as the conditional disability duration using Euler schemes (with step length 0.1 years) to solve the relevant Thiele equations. Aggregating the results, we obtain two set of predictions for each model proposal: total disability durations per group and total conditional disability durations per group.

Table 6.1 contains the mean absolute errors (MAE) and root mean square errors (RMSE) for the total durations per group and the total conditional durations per group. The results are consistent with our previous findings: the shrinkage approach outperforms the classic methods.

To conclude, we stress that this numerical example merely serves to illustrate that the simple shrinkage estimators proposed in this paper are a good addition to the toolbox of the actuarial practitioner. Often simple shrinkage estimation is not the best approach. Simulation of similar training and test datasets for which the simple shrinkage estimation procedure does not outperform the classic methods regarding MAE and RMSE supports this argument.

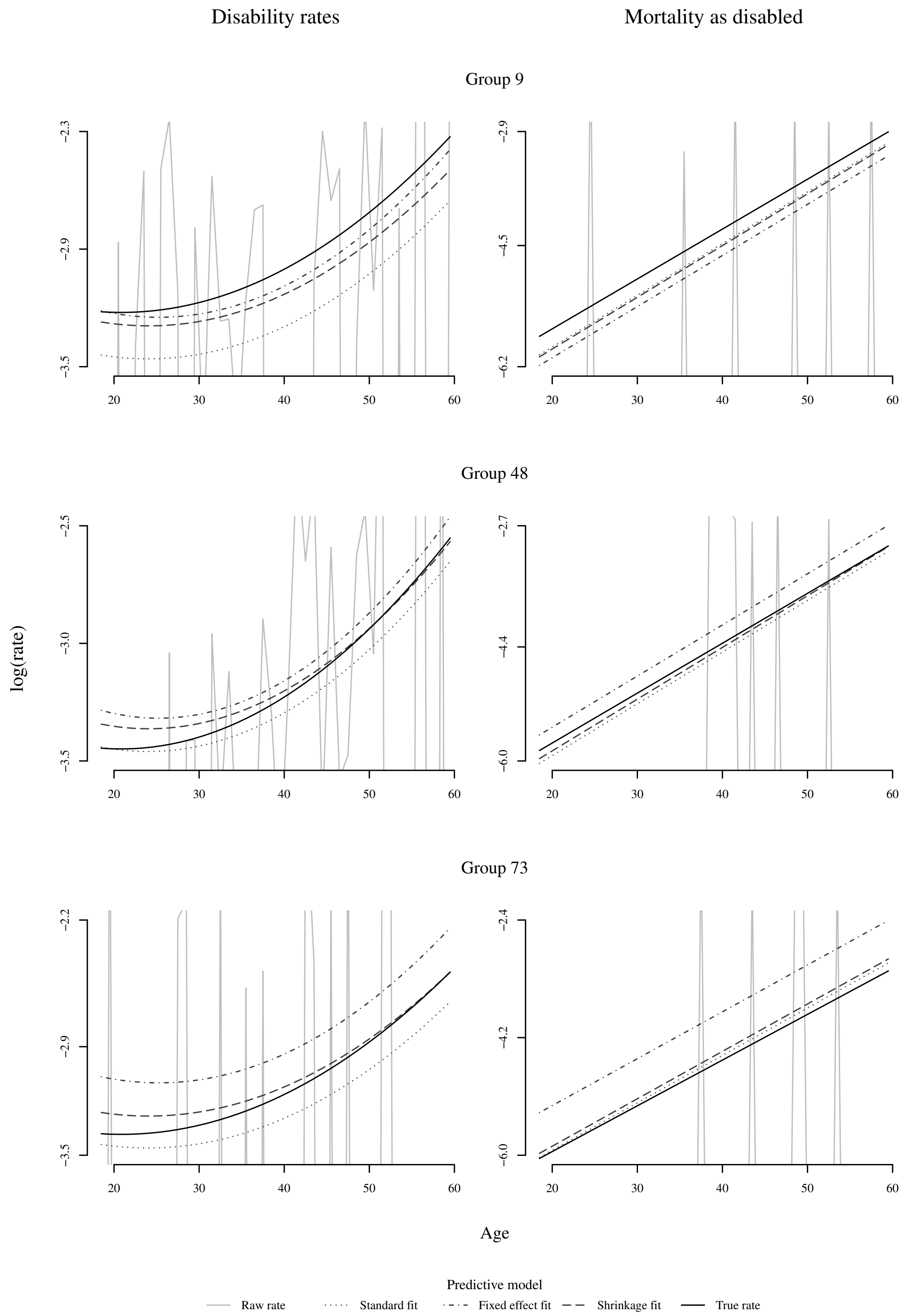

Figure 6.4: *Estimated disability rates (left) and estimated mortalities as disabled (right) with corresponding true rates and raw (observed) rates for three selected groups. Based on these three groups, the shrinkage estimation procedure appears to provide the best fit.*

6.5 Final remarks

We end this paper with a few comments related to hierarchical likelihoods, extensions to general truncation/censoring mechanisms, duration effects (so-called semi-Markov models), and time-dependent covariates, and related possibilities for future work.

Hierarchical likelihoods The empirical Bayes methods applied in Section 6.3 involve maximum likelihood estimation of the hyper-parameters and regression parameters using the marginal likelihood. For hierarchical generalized linear models, including multivariate negative binomial regressions, Lee and Nelder Lee and Nelder (1996) suggest estimation working with the hierarchical likelihood, the so-called h-likelihood, rather than the marginal likelihood. The h-likelihood is the full likelihood viewed not only as a function of the hyper-parameters and regression parameters, but also as a function of the latent variables. We stress that the characterization of models which yield a simpler (product) structure of the likelihoods applies in the same manner to the h-likelihood as the marginal likelihood, the latter being the object of interest in this paper. Consequently, an alternative class of shrinkage estimators is obtained by maximizing the resulting h-likelihoods rather than the marginal likelihoods.

Left-truncation and right-censoring In this paper, we assume the censoring mechanism to be given by deterministic right-censoring. As mentioned initially, this assumption is violated in practice as insured enter and exit the insurance portfolio randomly. Regarding model specification, we can work with a general truncation/censoring mechanism as long as it is conditionally independent and conditionally noninformative, see Andersen et al. (1988, Chapter IX). Furthermore, as the results of this paper are based on the structure of the likelihoods, they remain valid within such a setting.

At the end of Section 6.3, we stressed the importance of a careful analysis of the predictive performance of the simple shrinkage estimation procedure. When it comes to model-free consistent estimation of prediction errors, allowing for random truncation/censoring mechanisms is more troublesome, see also Gerds and Schu-

	Duration			*Conditional duration*		
	Standard	Fixed effect	Shrinkage	Standard	Fixed effect	Shrinkage
MAE	35.36	26.06	24.67	5.77	6.17	5.62
RMSE	47.71	35.42	34.64	7.59	8.14	7.51

Table 6.1: *Mean absolute errors (MAE) and root mean square errors (RMSE) for the total durations per group and total conditional durations per group. In all cases, the shrinkage approach leads to the smallest errors.*

macher (2006). This a general issue regarding model assessment in (multi-state) life insurance.

Duration effects: the semi-Markovian case The empirical Bayes and multivariate frailty extension can also be applied to semi-Markovian jump processes introduced in an actuarial context by Janssen (1966) and Hoem (1972). Using the terminology of Sokol (2015), when X is a semi-Markovian jump process with intensities, the corresponding multivariate counting process N has intensity processes

$$(0, \infty) \ni t \mapsto \lambda_{jk}(t) = \mathbb{1}_{(X(t-)=j)} \alpha_{jk}(t, U(t-))$$

where $U(t) = t - \sup\{s \in (0,t] : X(s-) \neq X(t)\}$ is the duration process measuring the amount of time spent by X in its current state and where α_{jk} are duration dependent transition intensities satisfying certain regularity conditions. The inclusion of duration effects does not fundamentally alter the (product) structure of the likelihood, so the characterizations given in Subsection 6.3.2 can be adapted to this setting. Furthermore, if the transition intensities are assumed to be piecewise constant as functions of time as well as duration, the link to Poisson regressions and multivariate negative binomial regressions remains present with duration dependent occurrences and exposures as sufficient statistics.

Time-dependent covariates and group compositions In this paper, we assume the covariates to be categorical and time-independent. In the same fashion, the group composition is assumed to be fixed in time contrary to practice where e.g. a natural group composition is given by employer or occupation which changes in time. Time-dependent covariates, and hence time-dependent group compositions, can be included in the framework by a suitable extension of the filtration without affecting the estimation procedures significantly, see e.g. Andersen et al. (1988) Section III.5. Each insured simply contributes to the occurrences and exposures of a specific group in the time period where he belongs to this group. In particular, the simple shrinkage estimation procedure is unchanged with a proper definition of occurrences and exposures taking the time-dependent covariates and time-dependent group composition into account.

Future work The natural next step is to evaluate the simple shrinkage estimation procedure on real disability insurance data taking the aforementioned extensions into account. This application ought to include an extended analysis of the predictive performance, which requires the derivation of consistent estimators for prediction errors under general truncation/censoring mechanisms.

Another direction for future work is to consider shrinkage estimation resulting from a class of priors that allows for dependency between latent variables within groups.

The class suggested by Norberg (1989a) is particularly interesting to practitioners as it produces interpretable expressions.

Acknowledgements

I would like to thank my supervisors Kristian Buchardt and Mogens Steffensen for valuable suggestions and fruitful discussions. I am also grateful to anonymous referees for comments and suggestions that improved the paper significantly.

Chapter 7

Tax- and expense-modified risk-minimization for insurance payment processes

This chapter contains the paper *Buchardt, Furrer, and Møller (2020).*

Abstract

We study the problem of determining risk-minimizing investment strategies for insurance payment processes in the presence of taxes and expenses. We consider the situation where taxes and expenses are paid continuously and symmetrically and introduce the concept of tax- and expense-modified risk-minimization. Risk-minimizing strategies in the presence of taxes and expenses are derived and linked to Galtchouk-Kunita-Watanabe decompositions associated with modified versions of the original payment processes. Furthermore, we show equivalence to an alternative approach involving an artificial market consisting of after-tax and after-expense assets, and we establish – in a certain sense – consistency with classic risk-minimization. Finally, a case study involving classic multi-state life insurance payments in combination with a bond market exemplifies the results.

Keywords: Quadratic hedging; Incomplete market; Market consistent valuation; Intrinsic value process; Galtchouk-Kunita-Watanabe decomposition

7.1 Introduction

According to a recent OECD report on taxation of funded private pension plans across different countries (OECD, 2015), taxes on pension fund returns are widespread.

Consequently, market consistent valuation of such insurance liabilities requires one to take into account the associated future tax payments, which are closely related to the investment strategy. Similarly, future expenses associated with the management of the insurance contract and investment strategy should be included in considerations about hedging and valuation. The necessity to take taxes and expenses into account is also reflected in the Solvency II regulation, see EIOPA (2009, Article 77-78) and EIOPA (2015, Article 28), and the forthcoming IFRS17 regulation, see IFRS (2017, Paragraph 34 and Paragraph B65(j)). It is our impression that a unified theory for market consistent valuation in the presence of taxes and expenses is yet to be developed, and accordingly, it is common among practitioners to take taxes and expenses into account via certain ad hoc adjustments of the forward interest rate curve, confer with Buchardt and Møller (2018).

In this paper, we consider quadratic hedging of insurance payment processes in the presence of taxes and expenses. We allow for idealized taxes and expenses which depend on the investment strategy and develop the concept of tax- and expense-modified risk-minimization. The taxes are defined as a fraction of the returns from the investment strategy, and the expenses are defined as a fraction of the value of the investment strategy, thus both are symmetrical and continuously paid. The primary idea is to introduce a tax- and expense-modified version of the so-called cost process and then minimize at any time the associated risk process, which is defined as the conditional expected value of the squared future tax- and expense-modified costs given the information currently available. The tax- and expense-modification of the costs reflects the impact of taxes and expenses on the time-value of money. As our main result, we show the existence and uniqueness of an optimal strategy and relate it to the Galtchouck-Kunita-Watanabe decomposition of the intrinsic value process associated with a tax- and expense-modified payment process.

In a complete market all contingent claims and liabilities can be hedged and prized uniquely via a no-arbitrage argument. The unique no-arbitrage price can be obtained by identifying a self-financing strategy, which replicates the liability perfectly from an initial investment and trading gains arising from trading in the underlying assets that are available in the market without adding additional capital. It then follows directly, that the price of the claim must coincide with the initial investment needed for this replication. Moreover, by following the replicating self-financing investment strategy, it is possible to eliminate all risk associated with the liability.

With an incomplete market, it is in general not possible to hedge a liability perfectly. Hence, it is not possible to price the liability directly from no-arbitrage arguments, and one cannot eliminate all risk associated with the liability. For example, this is relevant when studying the issue of valuation and hedging of insurance liabilities, see Møller (2002), who reviewed and studied various criteria from the literature on incomplete markets such as super-replication, quantile hedging, short-fall risk-minimization, risk-minimization, mean-variance hedging and utility indifference pricing.

Risk-mimization and mean-variance hedging are both so-called quadratic hedging principles since they involve a quadratic criterion for valuation and hedging. With mean-variance hedging, the idea is essentially to minimize the expected square value of the difference between the insurance liability and the terminal value of a self-financing strategy, i.e. to approximate the liability in an L_2-sense. With risk-minimization, one works directly under a martingale measure, where the discounted price processes are martingales and studies a larger class of strategies, so-called mean-self-financing strategies that allow for capital injections and withdrawals. In this setting, one introduces a cost process that keeps track of the additional investments needed for the investment strategy and the payment process of the liability. The risk-minimizing strategy is the strategy that minimizes the expected squared future costs at any time, and the initial investment required for this strategy can be viewed as a candidate for valuation of the liability.

Given a payment process in an incomplete market, it is well studied how to apply the quadratic hedging criterion of risk-minimization to find an optimal investment strategy and price the contract. The criterion of risk-minimization was originally proposed by Föllmer and Sondermann (1986) and was extended to insurance payment processes in Møller (2001); for an overview and general motivation for this choice of criterion, see Schweizer (2001).

Tax- and expense-modified risk-minimization differs from classic risk-minimization, in essence because a tax- and expense-modified savings account is used as numeraire. We show that tax- and expense-modified risk-minimization is consistent with classic risk-minimization in the sense that a subsequent application of classic risk-minimization confirms the investment strategy, thus not reducing the risk further.

In addition to the tax- and expense-modified risk-minimization approach, we also solve the problem by creating an artificial after-tax and after-expense market: The assets are constructed such that the returns are after payment of taxes and expenses. In this market, we are able to apply classic risk-minimization and thereby find an optimal investment strategy, which is essentially identical to the investment strategy from tax- and expense-modified risk-minimization. This is a consequence of the cost processes in the two approaches being in a certain sense identical.

Taxes on investment returns and expenses associated with the management of the insurance contract and investment strategy can be viewed as negative dividends. In that sense, the concept of tax- and expense-modified risk-minimization corresponds to a kind of risk-minimization in the presence of negative dividends. While the extension of risk-minimization to include transaction costs is studied in depth in the literature, see e.g. Lamberton, Pham, and Schweizer (1998) and Guasoni (2002), there does not seem to be a similar treatment in the literature of the case of dividends; an exception being Battauz (2003) on quadratic hedging in the presence of discrete stochastic dividends.

In Buchardt and Møller (2018), valuation of insurance payment processes in the presence of symmetric and continuously paid taxes and expenses is studied in complete markets. By identifying and explicitly constructing the inherent tax and expense payment processes and adding these to the existing insurance payment process, Buchardt and Møller (2018) were able to derive replicating strategies and determine the market value of the combined liability. In particular, by disregarding systematic insurance risk and implicitly dealing with unsystematic insurance risk via diversification, Buchardt and Møller (2018) were able to argue that current actuarial practice is prudent. In this paper, we essentially extend the results of Buchardt and Møller (2018) to allow for unsystematic as well as systematic insurance risk via an incomplete market approach.

Having determined the value of the combined liability, as well as the associated risk-minimizing investment strategy, it is interesting, as was done for complete markets in Buchardt and Møller (2018), to study the decomposition into benefit, tax, and expense parts. This is, however, together with extensions of the tax- and expense setup to include asymmetrical and discrete payments, postponed to future research.

The paper is structured in the following way. In Section 7.2, we give a brief review of the main results on risk-minimization for insurance payment processes. In Section 7.3, we study risk-minimization in the presence of symmetrical and continuously paid taxes and expenses. Section 7.4 concludes with a case study, where we consider classic multi-state life insurance payments in a bond market for a constant tax rate and expenses depending on the current state of the insurance contract(s).

7.2 Risk-minimization for insurance payment processes

In this section, we give a brief review of the main results on risk-minimization for insurance payment processes from Møller (2001), see also Schweizer (2001), before introducing taxes and expenses in the next section. Regarding the technical details and the necessary regularity conditions, we generally refer to Møller (2001) and Schweizer (2001) and the references therein.

Consider an arbitrage-free financial market consisting of $d+1$ traded assets with price processes S_0 and $(S_1, \dots, S_d)$ defined on a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\mathbb{F} = (\mathcal{F}(t))_{t\in[0,T]}$ satisfying the usual conditions with $\mathcal{F}(0)$ trivial. Here $T > 0$ is a fixed finite time horizon. We assume that S_0 is the savings account and that it is of the form

$$S_0(t) = \exp\biggl(\int_0^t r(u)\,\mathrm{d}u\biggr),$$

where r is the so-called short rate process.

All quantities are modeled directly under an equivalent martingale measure Q, such that the discounted price processes $S_j^* = S_j/S_0$ are Q-martingales. In general, results hold almost surely w.r.t. Q. We discuss the choice of equivalent martingale measure in the last paragraph of the present section.

We study an undiscounted *insurance payment process*, which is a stochastic process A describing the accumulated benefits less premiums associated with some insurance contract(s).

Let $S^* = (S_1^*, \ldots, S_d^*)$. Following Schweizer (1994, 2008), there exists a bounded, strictly increasing, predictable process B, null at 0, such that

$$\langle S_i^*, S_j^* \rangle \ll B$$

with $\langle \cdot \rangle$ denoting the predictable variation. Define matrix-valued predictable process σ_S by

$$\mathrm{d}\langle S^* \rangle = \sigma_S \,\mathrm{d}B. \tag{7.2.1}$$

Here each $\sigma_S(t)$ is a positive semidefinite symmetric $d \times d$-matrix. To ensure uniqueness of certain decompositions and optimal strategies in the sense that the amount invested in every asset is unique, we further assume that each $\sigma_S(t)$ is actually positive definite.

The following example illustrates the general concepts and also serves as the starting point for the case study in Section 7.4.

Example 7.2.1 (Classic multi-state model with investments in Vasicek bond market)**.** Especially in life insurance, one could be interested in the following more specific setting.

The financial market consists of two assets with price processes (S_0, S_1) driven by a stochastic short rate process r following a Vasicek model. In other words, r is an Ornstein-Uhlenbeck process satisfying the stochastic differential equation

$$\mathrm{d}r(t) = \kappa\,(\theta - r(t))\;\mathrm{d}t + \sigma\,\mathrm{d}W(t),$$

where κ, θ, and σ are positive constants and W is a standard Brownian motion under an equivalent martingale measure Q.

The development of an underlying life insurance contract (or multiple contracts) is described by the classic multi-state Markov model of Hoem (1969). Let Z be a

Markovian jump process with values in a finite set $\mathcal{J} = \{0, 1, \dots, J\}$ describing the state of the contract(s). The initial state of the contract(s) is taken to be 0 such that $Z(0) = 0$.

A multivariate counting process $N = (N_{jk})_{j,k\in\mathcal{J},k\neq j}$ is associated with the jump process Z by setting $N_{jk}(0) = 0$ and

$$N_{jk}(t) = \#\{s \in (0,t] : Z(s-) = j, Z(s) = k\}$$

for $t \in (0,T]$. The quantities $N_{jk}(t)$ can be interpreted as the number of transitions from state j to state k of the contract(s) within the time interval $[0,t]$.

We assume that Z and the financial market given by W are independent under Q, and we take the filtration $\mathbb{F}$ to be the Q-augmentation of the natural filtration of Z and W. The market is incomplete since the insurance risk is non-replicable.

In addition to the savings account S_0, the market contains a zero coupon bond with expiry at time $T > 0$. The price process is:

$$S_1(t) = \mathrm{E}^Q\left[\frac{S_0(t)}{S_0(T)}\,\middle|\,\mathcal{F}(t)\right] = \mathrm{E}^Q\left[e^{-\int_t^T r(s)\,\mathrm{d}s}\,\middle|\,\mathcal{F}(t)\right].$$

Since $d = 1$ and $\sigma > 0$, the condition of positive definiteness following (7.2.1) is trivially satisfied.

We assume there exist continuous functions $[0,T] \ni t \mapsto \mu_{jk}(t)$, $j,k \in \mathcal{J}$, $k \neq j$, such that Z has transition intensities μ completely characterizing the distribution of Z. It follows that the processes M_{jk} given by

$$M_{jk}(t) = N_{jk}(t) - \int_0^t 1_{\{Z(s-)=j\}}\mu_{jk}(s)\,\mathrm{d}s$$

are orthogonal martingales. Furthermore, each M_{jk} is also orthogonal to the discounted price process $S_1^* = S_1/S_0$.

We are interested in payment processes related to the development of the insurance contract(s). We denote these insurance payment processes by A^{b} and require they take the form

$$\mathrm{d}A^{\mathrm{b}}(t) = \sum_{j\in\mathcal{J}}\left(1_{\{Z(t)=j\}}b_j(t)\,\mathrm{d}t + \sum_{k:k\neq j} b_{jk}(t)\,\mathrm{d}N_{jk}(t)\right) \tag{7.2.2}$$

with $A^{\mathrm{b}}(0)$ some initial deterministic payment, and where b_j are deterministic sojourn payments and b_{jk} are deterministic transition payments all assumed measurable and bounded on bounded intervals. ∘

An *investment strategy* h is a $(d+1)$-dimensional process. Both the discounted price processes S_j^*, the insurance payment process A, the short rate process r, and the investment strategies h satisfy certain regularity conditions. The undiscounted *value process* V associated with h is defined by

$$V(h,t) = \sum_{j=0}^{d} h_j(t) S_j(t). \tag{7.2.3}$$

The value process measures the value of the investment strategy after the payments prescribed by A, i.e. $V(h,t)$ is the value of the investments after the payments $A(t)$ during $[0,t]$. We say that the investment strategy h is 0-*admissible* if the value at time T is 0, i.e. if $V(h,T) = 0$. The investment strategy is not assumed to be self-financing.

The undiscounted *cost process* C associated with h is defined by

$$C(h,t) = V(h,t) - \sum_{j=0}^{d} \int_0^t h_j(u)\, \mathrm{d}S_j(u) + A(t). \tag{7.2.4}$$

The value process measures the current value of the investment strategy h, and the cost process measures the accumulated costs associated with the investment strategy h and the insurance payment process A. The accumulated costs at time t are given by the current value of the investment portfolio, added past payments and reduced by realized trading gains.

Define the discounted value process by $V^* = V/S_0$, the discounted insurance payment process A^* by $A^*(0) = A(0)$ and

$$\mathrm{d}A^*(t) = S_0^{-1}(t)\, \mathrm{d}A(t),$$

and the discounted cost process C^* by $C^*(0) = C(0)$ and

$$\mathrm{d}C^*(t) = S_0^{-1}(t)\, \mathrm{d}C(t). \tag{7.2.5}$$

It follows that

$$\mathrm{d}C^*(h,t) = \mathrm{d}V^*(h,t) - \sum_{j=1}^{d} h_j(t)\, \mathrm{d}S_j^*(t) + \mathrm{d}A^*(t). \tag{7.2.6}$$

The *risk process* R associated with h and A is defined by

$$R(h,t) = \mathrm{E}^Q\Big[(C^*(h,T) - C^*(h,t))^2 \,\Big|\, \mathcal{F}(t) \Big]. \tag{7.2.7}$$

The process measures the quadratic risk under the measure Q associated with the future costs $(C^*(h,T) - C^*(h,t))$ given the information currently available. An investment strategy h is said to be *risk-minimizing* for A if it is 0-admissible and minimizes the risk process at any point in time.

Following Föllmer and Sondermann (1986) and Møller (2001), define the so-called *intrinsic value process* $\mathcal{V}^{A^*}$ associated with A^* by

$$\mathcal{V}^{A^*}(t) = \mathrm{E}^Q[A^*(T) \,|\, \mathcal{F}(t)] = A^*(t) + \mathrm{E}^Q\left[\int_t^T e^{-\int_0^s r(u)\,\mathrm{d}u}\,\mathrm{d}A(s)\,\middle|\,\mathcal{F}(t)\right]. \quad (7.2.8)$$

There exists a unique decomposition for $\mathcal{V}^{A^*}$ of the form

$$\mathcal{V}^{A^*}(t) = \mathcal{V}^{A^*}(0) + \sum_{j=1}^{d} \int_0^t h_j^{A^*}(u)\,\mathrm{d}S_j^*(u) + L^{A^*}(t), \quad (7.2.9)$$

where $h_1^{A^*}, \ldots, h_d^{A^*}$ satisfy certain regularity conditions, and where L^{A^*} is a zero-mean Q-martingale which is orthogonal to the discounted price processes S^*. The decomposition (7.2.9) is also known as the Galtchouk-Kunita-Watanabe decomposition.

Theorem 2.1 in Møller (2001) shows for the case $d = 1$ (an extension to the multi-dimensional case is possible; the assumption of positive definiteness following (7.2.1) ensures uniqueness) that there exists a unique risk-minimizing investment strategy h^* for A given by $h_j^* = h_j^{A^*}$ for $j = 1, \ldots, d$ and

$$h_0^*(t) = \mathcal{V}^{A^*}(t) - A^*(t) - \sum_{j=1}^{d} h_j^{A^*}(t) S_j^*(t). \quad (7.2.10)$$

Consequently, if one can explicitly write up the relevant Galtchouk-Kunita-Watanabe decomposition, this immediately yields an explicit risk-minimizing investment strategy for the insurance payment process.

The risk process associated with the the risk-minimizing investment strategy is

$$R(h^*, t) = \mathrm{E}^Q\left[\left(L^{A^*}(T) - L^{A^*}(t)\right)^2 \,\middle|\, \mathcal{F}(t)\right].$$

The value process associated with the risk-minimizing investment strategy is

$$\begin{aligned} V(h^*, t) &= e^{\int_0^t r(u)\,\mathrm{d}u}\left(\mathcal{V}^{A^*}(t) - A^*(t)\right) \\ &= \mathrm{E}^Q\left[\int_t^T e^{-\int_t^s r(u)\,\mathrm{d}u}\,\mathrm{d}A(s)\,\middle|\,\mathcal{F}(t)\right] \quad (7.2.11) \end{aligned}$$

due to (7.2.10), in particular, the value of the investments before any payments is

$$\begin{aligned} V(h^*, 0-) := V(h^*, 0) + A(0) &= \mathcal{V}^{A^*}(0) \\ &= A(0) + \mathrm{E}^Q\left[\int_0^T e^{-\int_0^s r(u)\,\mathrm{d}u}\,\mathrm{d}A(s)\right] \quad (7.2.12) \end{aligned}$$

by (7.2.8).

The criterion of risk-minimization must be defined in terms of a martingale measure Q, such that the discounted price processes are martingales, and it is important to note that the risk process (7.2.7) and hence also the risk-minimizing strategy depends on the choice of martingale measure Q. In the incomplete market, there are many different martingale measures, and this leaves open the question of how this measure should be chosen.

It would perhaps be more natural to introduce and minimize at any time t a risk process of the form (7.2.7), which involved an expected value with respect to the underlying probability measure P instead of a martingale measure Q. However, it follows from Schweizer (2001) that one cannot in general minimize at any time t functionals of the type (7.2.7) if the risk process is defined in terms of the probability measure P and if P is not a martingale measure. Instead, one can alternatively apply the criterion of local risk-minimization, which amounts to minimizing the risk in a more local manner. There is a close relation between the two criterions of (global) risk-minimization and local risk-minimization if the discounted price processes S^* are continuous. Indeed, if the discounted price processes are continuous, local risk-minimization with respect to P is equivalent to (global) risk-minimization with respect to a specific martingale measure, the so-called minimal martingale measure $\hat{Q}$. We refer to Schweizer (2008) for more details on this result. Thus, the minimal martingale measure is a natural candidate measure for global risk-minimization. However, the concept of risk-minimization can be used under any equivalent martingale measure, see also Møller (1998).

7.3 Risk-minimization in the presence of taxes and expenses

We extend the setting from Section 7.2 by including symmetrical taxes and expenses paid continuously. As in the previous section, we consider an insurance payment process A^{b} describing the accumulated benefits less premiums associated with some insurance contract(s).

We study two different approaches. First, we define after-tax and after-expense price processes directly from the underlying before-tax and before-expense price processes. These price processes are constructed exactly such that the return corresponds to the original return after taxes and expenses. In this setting with an artificial after-tax and after-expense market, we apply the criterion of risk-minimization directly. Second, we follow Buchardt and Møller (2018) and construct explicitly the payment processes associated with taxes and expenses and introduce the concept of tax- and expense-modified risk-minimization. The two approaches are conceptually different but are, as we unveil, mathematically equivalent in a specific sense which we explain later. The first approach using after-tax and after-expense price processes is detailed in Subsection 7.3.1, while Subsection 7.3.2 deals with tax- and expense-modified risk-minimization.

Following the second approach, where we have employed tax- and expense-modified risk-minimization, the risk-minimizing investment strategy leads to specific tax payments and expense payments. We investigate the following question: if another investor assumes these payments, would it using classic risk-minimization as in Section 7.2 employ a different optimal investment strategy than the original investor, who used the criterion of tax- and expense-modified risk-minimization? Unsurprisingly, the answer turns out negative; it is not possible to reduce the risk further. The investigation is presented in Subsection 7.3.3.

7.3.1 Risk-minimization in the after-tax and after-expense market

To model the taxes and expenses, we introduce a tax rate γ and an expense rate δ; both are adapted processes. We assume that γ takes values in $[0,1)$, has limits from the left, and is bounded away from 1, while δ is only assumed to be bounded (and measurable). The taxes are paid continuously at rate γ as a fraction of all returns (positive and negative) from the investment strategy, and the expenses are also paid continuously at rate δ but instead as a fraction of the value of the investment strategy.

The following example of taxes and expenses, which builds on the setup of Example 7.2.1, is the focal point of the case study in Section 7.4. It is introduced at this early stage to illustrate the general concepts.

Example 7.3.1 (Example 7.2.1 continued)**.** The tax rate $\gamma \in [0,1)$ is constant in time and deterministic. Thus the tax rate does not depend on the history of the insurance contract(s) or the history of the financial market. The expense rate δ is required to take the form

$$\delta(t) = \sum_{j \in \mathcal{J}} 1_{\{Z(t)=j\}} \delta_j(t)$$

for $t \in (0,T]$, where δ_j are continuous state-wise expense rates assumed deterministic. With Z describing the current state of the contract(s), this allows for modeling the administrative expenses more precisely, since various costs related to administration and investing typically depend on the state of the contract(s).

In general, the expense rate only depends on the history of the insurance contract(s) through the present state, and the expense rate does not depend on the history of the financial market. ○

The taxation and expense schemes introduced here are idealizations of real-life regimes. As an example, the Danish PAL-tax is a flat tax of 15.3 % on pension fund returns, but it is paid on a yearly basis (non-continuously) and asymmetrically: negative yearly returns lead to future tax deductions rather than negative tax payments. By setting $\gamma \equiv 0.153$ we obtain an idealization of the Danish taxation regime. In general, we consider our idealized approach a significant step towards understanding and handling a wide range of taxation and expense regimes.

Consider after-tax and after-expense price processes $\check{S}_j$ given by $\check{S}_j(0) = S_j(0)$ and

$$\mathrm{d}\check{S}_j(t) = \check{S}_j(t-)\left((1-\gamma(t-))\frac{\mathrm{d}S_j(t)}{S_j(t-)} - \delta(t)\,\mathrm{d}t\right), \tag{7.3.1}$$

for $j = 0, 1, \dots, d$, where S_j are the before-tax and before-expense price processes introduced in Section 7.2. In the following, we assume that the fractions $\check{S}_j/S_j$ are well-defined and that there exists suitably regular (strong) solutions to (7.3.1). We interpret the after-tax and after-expense price processes as price processes of an artificial after-tax and after-expense market; this is based on the following observation: Rewriting (7.3.1), we see that

$$\frac{\mathrm{d}\check{S}_j(t)}{\check{S}_j(t-)} = (1-\gamma(t-))\frac{\mathrm{d}S_j(t)}{S_j(t-)} - \delta(t)\,\mathrm{d}t,$$

which shows that the relative returns of the after-tax and after-expense assets are affine transformations of the relative returns of the original before-tax and before-expense assets. The relative returns are scaled with a factor $(1-\gamma)$ and reduced by δ. In other words, the returns (7.3.1) correspond to the returns obtained by an investor paying taxes and expenses according to the scheme described in the beginning of this subsection.

The after-tax and after-expense version of the savings account takes the form

$$\begin{aligned}\check{S}_0(t) &= \exp\left(\int_0^t ((1-\gamma(u))\,r(u) - \delta(u))\,\mathrm{d}u\right)\\ &= e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} S_0(t).\end{aligned} \tag{7.3.2}$$

This can be interpreted as using an artificial after-tax and after-expense short rate $((1-\gamma)r - \delta)$ rather than the original short rate r.

We now study the artificial after-tax and after-expense market $(\check{S}_0, \check{S}_1, \dots, \check{S}_d)$ within the setup of Section 7.2. Thus, we use the after-tax and after-expense savings account $\check{S}_0$ as numeraire, and we search for a risk-minimizing investment strategy $\check{h}$ for an insurance payment process A^{b} in the after-tax and after-expense market.

It can be shown that the discounted after-tax and after-expense price processes defined by $\check{S}_j^* = \check{S}_j/\check{S}_0$ have dynamics

$$\begin{aligned}
\mathrm{d}\check{S}_j^*(t) &= \check{S}_j^*(t-)\,(1-\gamma(t-))\left(\frac{\mathrm{d}S_j(t)}{S_j(t-)} - r(t)\,\mathrm{d}t\right) \\
&= \frac{\check{S}_j^*(t-)}{S_j(t-)} S_0(t)\,(1-\gamma(t-))\left((S_0(t))^{-1}\,\mathrm{d}S_j(t) - r(t)S_j^*(t-)\,\mathrm{d}t\right) \\
&= \frac{\check{S}_j^*(t-)}{S_j^*(t-)}\,(1-\gamma(t-))\,\mathrm{d}S_j^*(t). \qquad (7.3.3)
\end{aligned}$$

Note that $\check{S}_0$ rather than S_0 is used as numeraire in the definition of the discounted after-tax and after-expense price processes.

Since the discounted before-tax and before-expense price processes S_j^* are Q-martingales, it follows from (7.3.3) that the after-tax and after-expense price processes $\check{S}_j^*$ are Q-local martingales. In the following, we assume for simplicity that the after-tax and after-expense price processes actually are Q-martingales.

From (7.3.2) we see that

$$\mathrm{d}\check{S}_j^*(t) = \frac{\check{S}_j(t-)}{S_j(t-)} e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,(1-\gamma(t-))\,\mathrm{d}S_j^*(t), \qquad (7.3.4)$$

which relates the dynamics of the discounted after-tax price and after-expense price processes to the dynamics of the discounted before-tax and before-expense price processes.

In this setting, the discounted insurance payment process $\check{A}^{\mathrm{b},*}$ is given by $\check{A}^{\mathrm{b},*}(0) = A^{\mathrm{b}}(0)$ and

$$\mathrm{d}\check{A}^{\mathrm{b},*}(t) = \check{S}_0^{-1}(t)\,\mathrm{d}A^{\mathrm{b}}(t) = e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} S_0^{-1}(t)\,\mathrm{d}A^{\mathrm{b}}(t), \qquad (7.3.5)$$

the undiscounted and discounted value processes $\check{V}$ and $\check{V}^*$ associated with h are

$$\check{V}(h,t) = \sum_{j=0}^{d} h(t)\check{S}_j(t),$$

$$\check{V}^*(h,t) = \sum_{j=0}^{d} h(t)\check{S}_j^*(t),$$

the undiscounted and discounted cost processes $\check{C}$ and $\check{C}^*$ associated with h are

$$\check{C}(h,t) = \check{V}(h,t) - \sum_{j=1}^{d}\int_0^t h_j(u)\,\mathrm{d}\check{S}_j(u) + A^{\mathrm{b}}(t), \qquad (7.3.6)$$

$$\check{C}^*(\check{h},t) = \check{V}^*(h,t) - \sum_{j=1}^{d}\int_0^t h_j(u)\,\mathrm{d}\check{S}_j^*(u) + \check{A}^{\mathrm{b},*}(t), \qquad (7.3.7)$$

and the risk process associated with h is

$$\check{R}(h,t) = \mathrm{E}^Q\Big[\big(\check{C}^*(h,T) - \check{C}^*(h,t)\big)^2 \,\Big|\, \mathcal{F}(t)\Big].$$

It follows from the results reviewed in Section 7.2 that the risk-minimizing investment strategy for the insurance payment process A^{b} in the after-tax and after-expense market with numeraire $\check{S}_0$ can be expressed in terms of the Galtchouk-Kunita-Watanabe decomposition

$$\begin{aligned}\mathcal{V}^{\check{A}^{\mathrm{b},*}}(t) &= \mathrm{E}^Q\big[\check{A}^{\mathrm{b},*}(T)\,\big|\,\mathcal{F}(t)\big]\\ &= \mathcal{V}^{\check{A}^{\mathrm{b},*}}(0) + \sum_{j=1}^{d}\int_0^t \check{h}_j^{\check{A}^{\mathrm{b},*}}(u)\,\mathrm{d}\check{S}_j^*(u) + \check{L}^{\check{A}^{\mathrm{b},*}}(t),\end{aligned} \tag{7.3.8}$$

where the zero-mean martingale $\check{L}^{\check{A}^{\mathrm{b},*}}$ is orthogonal to the discounted after-tax and after-expense price processes $(\check{S}_1^*,\ldots,\check{S}_d^*)$.

With this notation in place, we can write up the following result, which is an immediate consequence of the results reviewed in Section 7.2.

Proposition 7.3.2. *The unique risk-minimizing investment strategy $\check{h}^*$ in the after-tax and after-expense market is given by*

$$\check{h}_j^*(t) = \check{h}_j^{\check{A}^{\mathrm{b},*}}(t) \tag{7.3.9}$$

for $j = 1,\ldots,d$ and

$$\check{h}_0^*(t) = \mathcal{V}^{\check{A}^{\mathrm{b},*}}(t) - \check{A}^{\mathrm{b},*}(t) - \sum_{j=1}^{d}\check{h}_j^*(t)\check{S}_j^*(t). \tag{7.3.10}$$

The associated value and risk processes are

$$\begin{aligned}\check{V}(\check{h}^*,t) &= \mathrm{E}^Q\left[\int_t^T e^{-\int_t^s((1-\gamma(u))r(u)-\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b}}(s)\,\middle|\,\mathcal{F}(t)\right], \qquad (7.3.11)\\ \check{R}(\check{h}^*,t) &= \mathrm{E}^Q\left[\left(\check{L}^{\check{A}^{\mathrm{b},*}}(T) - \check{L}^{\check{A}^{\mathrm{b},*}}(t)\right)^2\,\middle|\,\mathcal{F}(t)\right].\end{aligned}$$

From (7.3.11) we see that the current value of the investment strategy can be obtained as the conditional expected value of future payments, discounted with an artificial after-tax and after-expense short rate $((1-\gamma)r-\delta)$ rather than the original short rate r.

Recall that our interest lies in the before-tax and before-expense market rather than the artificial after-tax and after-expense market. We shall therefore now restate the risk-minimizing investment strategy in terms of quantities pertaining to the before-tax and before-expense market. Specifically, we want to identify an investment strategy h^* satisfying

$$V(h^*,t) = \sum_{j=0}^{d} h_j^*(t) S_j(t) = \sum_{j=0}^{d} \check{h}_j^*(t)\check{S}_j(t) = \check{V}(\check{h}^*,t).$$

An alternative Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\check{A}^{\mathrm{b},*}}$ is obtained by taking the discounted before-tax and before-expense price processes $(S_1^*,\ldots,S_d^*)$ as integrators, which yields

$$\mathcal{V}^{\check{A}^{\mathrm{b},*}}(t) = \mathcal{V}^{\check{A}^{\mathrm{b},*}}(0) + \sum_{j=1}^{d}\int_0^t h_j^{\check{A}^{\mathrm{b},*}}(u)\,\mathrm{d}S_j^*(u) + L^{\check{A}^{\mathrm{b},*}}(t), \tag{7.3.12}$$

where the zero-mean martingale $L^{\check{A}^{*,\mathrm{b}}}$ is orthogonal to the discounted before-tax and before-expense price processes $(S_1^*,\ldots,S_d^*)$. It moreover follows from (7.3.4) that

$$\mathrm{d}S_j^*(t) = \frac{S_j(t-)}{\check{S}_j(t-)} e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} \frac{1}{1-\gamma(t-)}\,\mathrm{d}\check{S}_j^*(t). \tag{7.3.13}$$

Because $L^{\check{A}^{*,\mathrm{b}}}$ is orthogonal to $(S_1^*,\ldots,S_d^*)$, we thus find that it is also orthorgonal to $(\check{S}_1^*,\ldots,\check{S}_d^*)$. By (7.3.13), we also find that

$$\begin{aligned}
&\mathcal{V}^{\check{A}^{\mathrm{b},*}}(t)\\
&= \mathcal{V}^{\check{A}^{\mathrm{b},*}}(0) + L^{\check{A}^{\mathrm{b},*}}(t)\\
&\quad + \sum_{j=1}^{d}\int_0^t h_j^{\check{A}^{\mathrm{b},*}}(u)\frac{S_j(u-)}{\check{S}_j(u-)} e^{-\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\frac{1}{1-\gamma(u-)}\,\mathrm{d}\check{S}_j^*(u).
\end{aligned} \tag{7.3.14}$$

Hence (7.3.14) is another Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\check{A}^{\mathrm{b},*}}$ w.r.t. $(\check{S}_1^*,\ldots,\check{S}_d^*)$. Uniqueness of the decomposition then implies $\check{L}^{\check{A}^{*,\mathrm{b}}} = L^{\check{A}^{*,\mathrm{b}}}$ and

$$\frac{\check{S}_j(t-)}{S_j(t-)}\check{h}_j^{\check{A}^{\mathrm{b},*}}(t) = \frac{1}{1-\gamma(t-)} e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} h_j^{\check{A}^{\mathrm{b},*}}(t) \tag{7.3.15}$$

for $j = 1,\ldots,d$.

While the risk-minimizing investment strategy of Proposition 7.3.2 pertains to the after-tax and after-expense market, it can be restated in terms of the before-tax and before-expense assets. Assume for a moment that the price processes are continuous, thus $S_j(t-) = S_j(t)$. An investment at time t of $\check{h}_j^*(t)$ in the after-tax and after-expense asset j corresponds to an investment of

$$h_j^*(t) = \frac{\check{S}_j(t)}{S_j(t)}\check{h}_j^*(t) \tag{7.3.16}$$

in the before-tax and before-expense asset j. It follows from (7.3.15) and (7.3.9) that

$$h_j^*(t) = \frac{1}{1-\gamma(t-)} e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} h_j^{\check{A}^{\mathrm{b},*}}(t), \tag{7.3.17}$$

for $j = 1, \ldots, d$, and from (7.3.10) and (7.3.2) that

$$\begin{aligned} h_0^*(t) &= e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} \left(\mathcal{V}^{\check{A}^{\mathrm{b},*}}(t) - \check{A}^{\mathrm{b},*}(t)\right) - \frac{\check{S}_0(t)}{S_0(t)} \sum_{j=1}^d \check{h}_j^*(t)\check{S}_j^*(t) \\ &= e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} \left(\mathcal{V}^{\check{A}^{\mathrm{b},*}}(t) - \check{A}^{\mathrm{b},*}(t)\right) - \sum_{j=1}^d h_j^*(t)S_j^*(t). \end{aligned} \tag{7.3.18}$$

Even if the price processes are not continuous, we can consider the investment strategy h^* given by

$$h_j^*(t) = \frac{\check{S}_j(t-)}{S_j(t-)} \check{h}_j^*(t) \tag{7.3.19}$$

for $j = 1, \ldots, d$ and

$$h_0^*(t) = \frac{\check{S}_0(t)}{S_0(t)} \check{h}_0^*(t) + (S_0(t))^{-1} \sum_{j=1}^d \check{h}_j^*(t)\check{S}_j(t) \left(1 - \frac{S_j(t)}{S_j(t-)} \frac{\check{S}_j(t-)}{\check{S}_j(t)}\right), \tag{7.3.20}$$

where the investment in the before-tax and before-expense savings account exactly has been determined such that the total value remains unchanged, i.e. such that

$$V(h^*, t) = \check{V}(\check{h}^*, t). \tag{7.3.21}$$

In particular, h^* is 0-admissible (pertaining to the before-tax and before-expense market). Furthermore, straightforward calculations show that h^* also in the discontinuous case satisfies (7.3.17) and (7.3.18). While h^* is 0-admissible, it is (in non-trivial cases) not risk-minimizing in the sense of minimizing the classic risk process R defined by (7.2.7) and based on the discounted cost process C^* given by (7.2.5). In the following subsection, we arrive at the same investment strategy by explicitly constructing the payment processes associated with taxes and expenses and applying the new concept of tax- and expense-modified risk-minimization rather than classic risk-minimization.

7.3.2 Tax- and expense-modified risk-minimization

In this subsection, we consider an alternative approach and construct explicitly payment processes related to taxes and expenses. Let γ and δ be the tax and expense rates introduced in Subsection 7.3.1. Since the taxes and expenses lead to payments that depend on the investment strategy and the investment returns, we introduce two additional payment processes $A^{\mathrm{tax}}(h)$ and $A^{\mathrm{e}}(h)$ for taxes and expenses, respectively, defined by $A^{\mathrm{tax}}(h,0) = 0$, $A^{\mathrm{e}}(h,0) = 0$, and

$$\mathrm{d}A^{\mathrm{tax}}(h,t) = \gamma(t-) \sum_{j=0}^{d} h_j(t)\,\mathrm{d}S_j(t), \tag{7.3.22}$$

$$\mathrm{d}A^{\mathrm{e}}(h,t) = \delta(t) V(h,t)\,\mathrm{d}t. \tag{7.3.23}$$

We note that the taxes are symmetric in the sense that positive investment returns lead to a tax payment, whereas negative investment returns lead to a tax income (negative payment). We can interpret the taxes and expenses as negative dividends, by introducing dividend processes D_j given by $D_j(0) = 0$ and

$$\mathrm{d}D_j(t) = -\gamma(t-)\,\mathrm{d}S_j(t) - \delta(t) S_j(t)\,\mathrm{d}t$$

for $j = 0, \ldots, d$, when

$$\mathrm{d}A^{\mathrm{tax}}(h,t) + \mathrm{d}A^{\mathrm{e}}(h,t) = -\sum_{j=0}^{d} h_j(t)\,\mathrm{d}D_j(t).$$

The traded assets with price processes $(S_0, S_1, \ldots, S_d)$ are then to be seen as trading ex dividend (or before taxes and expenses). In comparison, the artificial market with price processes $(\check{S}_0, \check{S}_1, \ldots, \check{S}_d)$ of Subsection 7.3.1 was given the interpretation of being after taxes and expenses. As mentioned in Section 7.1, risk-minimization in the presence of dividends appears to be a rather unexplored area of research; for a general introduction to dividends in continuous time we refer to Björk (2009, Chapter 16).

We are interested in the problem of determining risk-minimizing investment strategies for the combined payments consisting of the three payment processes A^{b}, $A^{\mathrm{tax}}(h)$, and $A^{\mathrm{e}}(h)$. Thus, we define the undiscounted cost process in the presence of taxes and expenses as the cost process of the total payments, which depend on the choice of our object of interest, namely the investment strategy h. In this sense, we are facing a fixed-point problem.

Definition 7.3.3. *The undiscounted cost process C in the presence of taxes and expenses associated with an investment strategy h and an insurance payment process A^{b} is defined by*

$$C(h,t) = V(h,t) - \sum_{j=0}^{d} \int_0^t h_j(u)\,\mathrm{d}S_j(u) + A^{\mathrm{b}}(t) + A^{\mathrm{tax}}(h,t) + A^{\mathrm{e}}(h,t), \tag{7.3.24}$$

where $A^{\mathrm{tax}}(h)$ and $A^{\mathrm{e}}(h)$ are defined by (7.3.22) and (7.3.23), respectively.

We see that $C(h,t)$ comprises the accumulated costs during $[0,t]$ including the payments $A^{\mathrm{b}}(t)$, $A^{\mathrm{tax}}(h,t)$, and $A^{\mathrm{e}}(h,t)$. Therefore, the value process at time t, i.e.

$V(h,t)$, should, in a similar fashion to previously, be interpreted as the value of the portfolio h held at time t *after* all payments during $[0,t]$, *including taxes and expenses.*

The discounted cost process C^* is defined from the undiscounted cost process via (7.2.5), i.e.

$$C^*(h,t) = C(h,0) + \int_0^t S_0^{-1}(u)\,\mathrm{d}C(h,u). \tag{7.3.25}$$

We now introduce the following definitions of tax- and expense-modified value, cost, and risk processes, respectively.

Definition 7.3.4. *The tax- and expense-modified value and cost processes are defined via*

$$\tilde{V}(h,t) = e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} V^*(h,t), \tag{7.3.26}$$

$$\tilde{C}(h,t) = C^*(h,0) + \int_0^t e^{\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}C^*(h,u), \tag{7.3.27}$$

where C^* is defined by (7.3.25)*, and where C is defined by* (7.3.24)*. A strategy is said to be risk-minimizing for A^{b} in the presence of taxes and expenses if it is* 0*-admissible and minimizes for all $t \in [0,T]$ the tax- and expense-modified risk process $\tilde{R}$ defined by*

$$\tilde{R}(h,t) = \mathrm{E}^Q\left[\left(\tilde{C}(h,T) - \tilde{C}(h,t)\right)^2 \middle| \mathcal{F}(t)\right]. \tag{7.3.28}$$

Note that the tax- and expense-modified quantities correspond to the usual discounted quantities but using as numeraire the after-tax and after-expense savings account rather than the before-tax and before-expense savings account.

By straightforward calculations, we find that

$$\begin{aligned}
&e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}C^*(h,t) \\
&\qquad = \mathrm{d}\Big(e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} V^*(h,t)\Big) + e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b},*}(t) \\
&\qquad\quad - \sum_{j=1}^d h_j(t)\,(1-\gamma(t-))\,e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}S_j^*(t).
\end{aligned}$$

It follows from the definition given by (7.3.27) and the above calculations that

$$\begin{aligned}
\mathrm{d}\tilde{C}(h,t) = {} & \mathrm{d}\tilde{V}(h,t) - \sum_{j=1}^d h_j(t)\,(1-\gamma(t-))\,e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}S_j^*(t) \\
& + e^{\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b},*}(t).
\end{aligned} \tag{7.3.29}$$

Thus, the dynamics of the tax- and expense-modified cost process $\tilde{C}$ has a structure which is similar to the dynamics for the discounted cost process in the traditional setting, compare with (7.2.6). This observation enables us to use similar techniques as in the classic setting without taxes and expenses for determining risk-minimizing investment strategies. To do this, we first define a tax- and expense-modified version $\tilde{A}^{\mathrm{b}}$ of the discounted insurance payment process $A^{\mathrm{b},*}$ via

$$\tilde{A}^{\mathrm{b}}(t) = A^{\mathrm{b},*}(0) + \int_0^t e^{\int_0^s (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b},*}(s).$$

Note that the tax- and expense-modified insurance payment process corresponds to the usual discounted insurance payment process but with the after-tax and after-expense savings account rather than the before-tax and before-expense savings account as numeraire, i.e.

$$\tilde{A}^{\mathrm{b}} = \breve{A}^{\mathrm{b},*}, \tag{7.3.30}$$

where $\breve{A}^{\mathrm{b},*}$ is given by (7.3.5). The new notation is solely to stress the change in interpretation compared to the Subsection 7.3.1.

We again define the intrinsic value process associated with $\tilde{A}^{\mathrm{b}}$ as the Q-martingale

$$\begin{aligned}
\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) &= \mathrm{E}^{Q}\Big[\tilde{A}^{\mathrm{b}}(T)\Big|\,\mathcal{F}(t)\Big] && (7.3.31)\\
&= \tilde{A}^{\mathrm{b}}(t) + \mathrm{E}^{Q}\left[\int_t^T e^{-\int_0^s r(u)\,\mathrm{d}u} e^{\int_0^s (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b}}(s)\,\middle|\,\mathcal{F}(t)\right], && (7.3.32)
\end{aligned}$$

and (re)write its Galtchouk-Kunita-Watanabe decomposition as

$$\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) = \mathcal{V}^{\tilde{A}^{\mathrm{b}}}(0) + \sum_{j=1}^{d}\int_0^t h_j^{\tilde{A}^{\mathrm{b}}}(u)\,\mathrm{d}S_j^*(u) + L^{\tilde{A}^{\mathrm{b}}}(t). \tag{7.3.33}$$

Due to uniqueness of the decomposition, we have $\mathcal{V}^{\tilde{A}^{\mathrm{b}}} = \mathcal{V}^{\breve{A}^{\mathrm{b},*}}$, $L^{\tilde{A}^{\mathrm{b}}} = \breve{L}^{\breve{A}^{\mathrm{b},*}}$, and $h_j^{\tilde{A}^{\mathrm{b}}} = h_j^{\breve{A}^{\mathrm{b},*}}$ for $j = 1, \ldots, d$, confer also with (7.3.30) and (7.3.12).

The following theorem contains the main result of the paper.

Theorem 7.3.5. *There exists a unique risk-minimizing investment strategy $\tilde{h}$ for A^{b} in the presence of taxes and expenses given by*

$$\tilde{h}_j(t) = \frac{1}{1-\gamma(t-)} e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u} h_j^{\tilde{A}^{\mathrm{b}}}(t), \tag{7.3.34}$$

for $j = 1, \ldots, d$ and

$$\tilde{h}_0(t) = e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\Big(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\Big) - \sum_{j=1}^{d}\tilde{h}_j(t)S_j^*(t). \tag{7.3.35}$$

The associated risk process is given by

$$\tilde{R}(\tilde{h},t) = \mathrm{E}^Q\left[\left(L^{\tilde{A}^{\mathrm{b}}}(T) - L^{\tilde{A}^{\mathrm{b}}}(t)\right)^2 \middle| \mathcal{F}(t)\right], \tag{7.3.36}$$

and the associated value process is

$$V(\tilde{h},t) = \mathrm{E}^Q\left[\int_t^T e^{-\int_t^s ((1-\gamma(u))r(u)-\delta(u))\,\mathrm{d}u}\,\mathrm{d}A^{\mathrm{b}}(s) \middle| \mathcal{F}(t)\right]. \tag{7.3.37}$$

The proof of the result is presented below. First, we give an interpretation of the result and relate it to the risk-minimizing investment strategy in the after-tax and after-expense market, confer with Proposition 7.3.2 and the discussion thereafter.

By comparing (7.3.33) and (7.2.9), we see that the quantity $h_j^{\tilde{A}^b}(t)$ can be interpreted as the number of assets j at time t in a risk-minimizing investment strategy for the modified payment process $\tilde{A}^{\mathrm{b}}$ in the classic setting without taxes and expenses, see Section 7.2. The solution of Theorem 7.3.5 is a modification of this strategy which adjusts for the taxes and expenses via the factors $(1-\gamma(t-))^{-1}$ and $e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}$.

From (7.3.37) we see that the current value of the investment strategy can be obtained as the conditional expected value of future payments given the information currently available but discounted with a modified short rate $((1-\gamma)r-\delta)$ rather than the original short rate r.

The modified discount factor corresponds to using the after-tax and after-expense savings account rather than the before-tax and before-expense savings account as the reference for the time-value of money. Moreover, the value agrees with the value obtained in Subsection 7.3.1 for the after-tax and after-expense market, confer with (7.3.11), i.e.

$$V(\tilde{h},t) = \check{V}(\check{h}^*,t)$$

with $\check{h}^*$ given by (7.3.9) and (7.3.10). Actually, we see that $\tilde{h} = h^*$ with h^* given by (7.3.19) and (7.3.20). Furthermore, since $L^{\tilde{A}^{\mathrm{b}}} = \check{L}^{\check{A}^{\mathrm{b},*}}$, the risk processes associated with the risk-minimizing strategies of course also agree; in other words, the residual risks are identical. This proves the relation between after-tax and after-expense risk-minimization and tax- and expense-modified risk-minimization alluded to at the end of Subsection 7.3.1. In other words, strategies resulting from the two conceptually different approaches are mathematically equivalent in the sense that the sums invested in each asset are equal.

The proof of Theorem 7.3.5 can in principle be based on Proposition 7.3.2. Straightforward calculations show that the modified and discounted cost processes $\tilde{C}$ and $\check{C}^*$, respectively, are directly related by adjusting the investment strategies according to the mappings given by (7.3.19) and (7.3.20). Thus by essentially combining (7.3.15) and (7.3.12) with the identity (7.3.30), the proof can be established. To better reveal what is going on behind the scenes, we provide a direct proof.

Proof of Theorem 7.3.5. The result is proven by determining the quantity (7.3.28) and minimizing it. Since by definition $A^{\text{tax}}(h,0) = 0$ and $A^{\text{e}}(h,0) = 0$, we see that $\tilde{C}(h,0) = \tilde{V}(h,0) + \tilde{A}^{\text{b}}(0)$. It now follows from (7.3.29) and the definition of the modified cost process that

$$\begin{aligned}\tilde{C}(h,t) &= \tilde{V}(h,t) + \tilde{A}^{\text{b}}(t)\\ &\quad - \sum_{j=1}^{d} \int_0^t (1-\gamma(u-))\, h_j(u) e^{\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\, \mathrm{d}S_j^*(u).\end{aligned} \tag{7.3.38}$$

For 0-admissible strategies $\tilde{h}$ we have that $V(\tilde{h},T) = 0$ and hence also $\tilde{V}(\tilde{h},T) = 0$. Moreover, (7.3.33) implies that

$$\tilde{A}^{\text{b}}(T) = \mathcal{V}^{\tilde{A}^{\text{b}}}(T) = \mathcal{V}^{\tilde{A}^{\text{b}}}(0) + \sum_{j=1}^{d} \int_0^T h_j^{\tilde{A}^{\text{b}}}(u)\, \mathrm{d}S_j^*(u) + L^{\tilde{A}^{\text{b}}}(T). \tag{7.3.39}$$

Now insert (7.3.39) into (7.3.38) with $t = T$ and use $\tilde{V}(\tilde{h},T) = 0$ to see that

$$\begin{aligned}\tilde{C}(\tilde{h},T) &= \mathcal{V}^{\tilde{A}^{\text{b}}}(0) + L^{\tilde{A}^{\text{b}}}(T)\\ &\quad + \sum_{j=1}^{d} \int_0^T \Big(h_j^{\tilde{A}^{\text{b}}}(u) - (1-\gamma(u-))\, \tilde{h}_j(u) e^{\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau} \Big)\, \mathrm{d}S_j^*(u)\\ &= \mathcal{V}^{\tilde{A}^{\text{b}}}(t) + \Big(L^{\tilde{A}^{\text{b}}}(T) - L^{\tilde{A}^{\text{b}}}(t) \Big)\\ &\quad - \sum_{j=1}^{d} \int_0^t (1-\gamma(u-))\, \tilde{h}_j(u) e^{\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\, \mathrm{d}S_j^*(u)\\ &\quad + \sum_{j=1}^{d} \int_t^T \Big(h_j^{\tilde{A}^{\text{b}}}(u) - (1-\gamma(u-))\, \tilde{h}_j(u) e^{\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau} \Big)\, \mathrm{d}S_j^*(u),\end{aligned}$$

where the last equality follows from (7.3.33). By combining the expressions for $\tilde{C}(\tilde{h},T)$ and $\tilde{C}(\tilde{h},t)$ and by using the orthogonality of the Q-martingales S_j^* and $L^{\tilde{A}^{\text{b}}}$, we find that

$$\tilde{R}(\tilde{h},t) = R_1(\tilde{h},t) + \mathrm{E}^Q\Big[\Big(L^{\tilde{A}^{\text{b}}}(T) - L^{\tilde{A}^{\text{b}}}(t) \Big)^2 \,\Big|\, \mathcal{F}(t) \Big] + R_2(\tilde{h},t)$$

with R_1 and R_2 given by

$$R_1(\tilde{h},t) = \left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{V}(\tilde{h},t) - \tilde{A}^{\mathrm{b}}(t)\right)^2$$

$$R_2(\tilde{h},t)$$

$$= \mathrm{E}^Q\left[\left(\sum_{j=1}^{d}\int_t^T\left(h_j^{\tilde{A}^{\mathrm{b}}}(u) - (1-\gamma(u-))\,\tilde{h}_j(u)e^{\int_0^u(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\right)\mathrm{d}S_j^*(u)\right)^2\middle|\,\mathcal{F}(t)\right].$$

The terms R_1 and R_2 can now be eliminated as follows. First, the term R_2 is eliminated by e.g. choosing $(\tilde{h}_1,\ldots,\tilde{h}_d)$ according to (7.3.34). By examining R_1, we realize that this term next can be eliminated if and only if

$$e^{\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}V^*(\tilde{h},t) = \tilde{V}(\tilde{h},t) = \mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t),$$

i.e. if and only if

$$V^*(\tilde{h},t) = e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right),$$

which is then obtained by choosing $\tilde{h}_0$ uniquely according to (7.3.35). This shows that $\tilde{h}$ given by (7.3.34) and (7.3.35) is a risk-minimizing investment strategy for A^{b} in the presence of taxes and expenses, and, further, establishes (7.3.36) and (7.3.37).

To prove uniqueness, it remains to be shown that each $\tilde{h}_j$ for $j=1,\ldots,d$ is uniquely determined. First note that by Itô isometry and (7.2.1),

$$\begin{aligned}R_2(\tilde{h},0) &= \sum_{i=1}^{d}\sum_{j=1}^{d}\mathrm{E}^Q\left[\int_0^T\alpha_i(t)\alpha_j(t)\,\mathrm{d}\langle S_i^*,S_j^*\rangle(t)\right]\\ &= \mathrm{E}^Q\left[\int_0^T\alpha^{\mathrm{tr}}(t)\sigma_S(t)\alpha(t)\,\mathrm{d}B(t)\right],\end{aligned}$$

where $\alpha=(\alpha_1,\ldots,\alpha_d)$ is given by

$$\alpha_j(t) = h_j^{\tilde{A}^{\mathrm{b}}}(t) - (1-\gamma(t-))\,\tilde{h}_j(t)e^{\int_0^t(\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}$$

for $j=1,\ldots,d$. Recall that each $\sigma_S(t)$ is assumed positive definite and that B is null at 0 and strictly increasing. It follows that $R_2(\tilde{h},0)=0$ if and only if each α_j is zero, i.e. if and only if $(\tilde{h}_1,\ldots,\tilde{h}_d)$ is chosen according to (7.3.34). This proves uniqueness of the risk-minimizing investment strategy. □

In this subsection, we have introduced the concept of tax- and expense-modified risk-minimization and solved the corresponding minimization problem based on the tax-and expense-modified risk process $\tilde{R}$ defined by (7.3.28), which involves the tax- and expense-modified cost process $\tilde{C}$ defined by (7.3.27) using a modified discount factor $(1-\gamma)r-\delta$.

We should like to stress that we have not minimized the risk process

$$(h,t) \mapsto \mathrm{E}^Q\Big[(C^*(h,T) - C^*(h,t))^2 \Big| \mathcal{F}(t)\Big] \tag{7.3.40}$$

with C^* given by (7.3.25) and (7.3.24).

Correspondingly, the solutions we provide here do not allow us to draw any conclusions concerning this particular minimization problem.

As previously mentioned, the modified discount factor corresponds to using the after-tax and after-expense savings account rather than the before-tax and before-expensive savings account as the reference for the time value of money. In other words, modified risk-minimization takes into account the fact that taxes and expenses essentially impact the time-value of money. In our opinion, modified risk-minimization is therefore more instinctive than what could have been a first Ansatz, namely simply minimizing (7.3.40).

7.3.3 Two-step risk-minimization

In this subsection, we study two-step risk-minimization in the following sense. Assume that the original investor, say an insurer, adopts the risk-minimizing investment strategy $\tilde{h}$ in the presence of taxes and expenses given by Theorem 7.3.5. Then it faces the tax specific payments $A^{\mathrm{tax}}(\tilde{h})$ and expense payments $A^{\mathrm{e}}(\tilde{h})$ in addition to the original insurance payments A^{b}. We consider the scenario where a systemic investor, e.g. a re-insurer, assumes the payments and faces the problem of classic risk-minimization (in the absence of taxes and expenses) within the setup of Section 7.2. The relevant classic insurance payment process A is thus given by

$$A(t) = A^{\mathrm{b}}(t) + A^{\mathrm{tax}}(\tilde{h},t) + A^{\mathrm{e}}(\tilde{h},t), \tag{7.3.41}$$

where $\tilde{h}$ is determined by Theorem 7.3.5 and thus fixed. To determine the classic risk-minimizing investment strategy for A, we now study the intrinsic value process $\mathcal{V}^{A^*}$ associated with A^*. The following lemma is key.

Lemma 7.3.6. *With A given by* (7.3.41), *the corresponding intrinsic value process $\mathcal{V}^{A^*}$ has the following Galtchouk-Kunita-Watanabe decomposition:*

$$\mathcal{V}^{A^*}(t) = \mathcal{V}^{A^*}(0) + \sum_{j=1}^d \int_0^t \tilde{h}_j(u)\,\mathrm{d}S_j^*(u) + \int_0^t e^{-\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}L^{\tilde{A}^{\mathrm{b}}}(u),$$

with $\tilde{h}$ the risk-minimizing investment strategy of A^{b} in the presence of taxes and expenses given by Theorem 7.3.5 and with $L^{\tilde{A}^{\mathrm{b}}}$ as in the Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\tilde{A}^{\mathrm{b}}}$, confer with (7.3.33)*. Furthermore,*

$$\mathcal{V}^{A^*}(t) = A^*(t) + e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\Big(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\Big).$$

Proof. By straightforward calculations we obtain the following expression:

$$\begin{aligned}
A^*(t) &= \int_0^t e^{-\int_0^u r(\tau)\,\mathrm{d}\tau}\,\mathrm{d}A(u) \\
&= A^{\mathrm{b},*}(t) + \sum_{j=1}^d \int_0^t \gamma(u-)\tilde{h}_j(u)\,\mathrm{d}S_j^*(u) \\
&\qquad + \int_0^t (\gamma(u)r(u)+\delta(u))\,e^{-\int_0^u r(\tau)\,\mathrm{d}\tau} V(\tilde{h},u)\,\mathrm{d}u.
\end{aligned}$$

Because the discounted price processes are Q-martingales, it follows that

$$\begin{aligned}
\mathcal{V}^{A^*}(t) &= \mathrm{E}^Q[\,A^*(T)\,|\,\mathcal{F}(t)] \\
&= \mathcal{V}^{A^{\mathrm{b},*}}(t) + A^{\mathrm{tax},*}(\tilde{h},t) + A^{\mathrm{e},*}(\tilde{h},t) \\
&+ \mathrm{E}^Q\left[\int_t^T (\gamma(u)r(u)+\delta(u))\,e^{-\int_0^u r(\tau)\,\mathrm{d}\tau} V(\tilde{h},u)\,\mathrm{d}u \,\middle|\, \mathcal{F}(t)\right].
\end{aligned}$$

From (7.3.37),

$$V(\tilde{h},u) = \mathrm{E}^Q\left[\int_u^T e^{-\int_u^s((1-\gamma(\tau))r(\tau)-\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}A^{\mathrm{b}}(s) \,\middle|\, \mathcal{F}(u)\right].$$

By the law of iterated expectations and by interchanging the order of integration, simple manipulations yield

$$\begin{aligned}
&\mathrm{E}^Q\left[\int_t^T (\gamma(u)r(u)+\delta(u))\,e^{-\int_0^u r(\tau)\,\mathrm{d}\tau} V(\tilde{h},u)\,\mathrm{d}u \,\middle|\, \mathcal{F}(t)\right] \\
&= \mathrm{E}^Q\left[\int_t^T e^{-\int_0^s r(\tau)\,\mathrm{d}\tau} \int_t^s (\gamma(u)r(u)+\delta(u))\,e^{\int_u^s(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}u\,\mathrm{d}A^{\mathrm{b}}(s) \,\middle|\, \mathcal{F}(t)\right] \\
&= \mathrm{E}^Q\left[\int_t^T e^{-\int_0^s r(\tau)\,\mathrm{d}\tau} \left(e^{\int_t^s(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau} - 1\right)\mathrm{d}A^{\mathrm{b}}(s) \,\middle|\, \mathcal{F}(t)\right] \\
&= e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right) - \left(\mathcal{V}^{A^{\mathrm{b},*}}(t) - A^{\mathrm{b},*}(t)\right),
\end{aligned}$$

see also (7.2.8) and (7.3.32). Collecting everything, we obtain

$$\begin{aligned}
\mathcal{V}^{A^*}(t) &= \mathcal{V}^{A^{\mathrm{b},*}}(t) + A^{\mathrm{tax},*}(\tilde{h},t) + A^{\mathrm{e},*}(\tilde{h},t) \\
&\quad + e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right) - \left(\mathcal{V}^{A^{\mathrm{b},*}}(t) - A^{\mathrm{b},*}(t)\right) \\
&\quad = A^{\mathrm{b},*}(t) + A^{\mathrm{tax},*}(\tilde{h},t) + A^{\mathrm{e},*}(\tilde{h},t) + e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right) \\
&\quad = A^*(t) + e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right),
\end{aligned}$$

as desired. Now using integration by parts and the definition of $\tilde{A}^{\mathrm{b}}$, we find that

$$\mathrm{d}\mathcal{V}^{A^*}(t) = \sum_{j=1}^d \gamma(t-)\tilde{h}_j(t)\,\mathrm{d}S_j^*(t) + e^{-\int_0^t(\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t).$$

From the Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\tilde{A}^{\mathrm{b}}}$, confer with (7.3.33), and the identity (7.3.34), it follows that

$$\begin{aligned}
\mathrm{d}\mathcal{V}^{A^*}(t) &= \sum_{j=1}^{d} \gamma(t-)\tilde{h}_j(t)\,\mathrm{d}S_j^*(t) + \sum_{j=1}^{d}(1-\gamma(t-))\tilde{h}_j(t)\,\mathrm{d}S_j^*(t) \\
&\quad + e^{-\int_0^t (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}L^{\tilde{A}^{\mathrm{b}}}(t) \\
&= \sum_{j=1}^{d} \tilde{h}_j(t)\,\mathrm{d}S_j^*(t) + e^{-\int_0^t (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}L^{\tilde{A}^{\mathrm{b}}}(t).
\end{aligned}$$

Note that the final term is a zero-mean Q-martingale orthogonal to the discounted price processes, because this is the case for $L^{\tilde{A}^{\mathrm{b}}}$. We conclude that $\mathcal{V}^{A^*}$ has Galtchouk-Kunita-Watanabe decomposition given by

$$\mathcal{V}^{A^*}(t) = \mathcal{V}^{A^*}(0) + \sum_{j=1}^{d}\int_0^t \tilde{h}_j(u)\,\mathrm{d}S_j^*(u) + \int_0^t e^{-\int_0^u (\gamma(\tau)r(\tau)+\delta(\tau))\,\mathrm{d}\tau}\,\mathrm{d}L^{\tilde{A}^b}(u),$$

as desired. □

Combining Lemma 7.3.6 with the classic results reviewed in Section 7.2, we obtain the main result of this subsection:

Proposition 7.3.7. *With A given by* (7.3.41) *there exists a unique classic risk-minimizing investment strategy for A, and this investment strategy is identical to the unique risk-minimizing investment strategy for A^{b} in the presence of taxes and expenses.*

Proof. It follows from the results reviewed in Section 7.2 that there exists a unique classic risk-minimizing investment strategy h^* for A given by $h_j^* = \tilde{h}_j$ for $j = 1, \ldots, d$, due to Lemma 7.3.6, and

$$h_0^*(t) = \mathcal{V}^{A^*}(t) - A^*(t) - \sum_{j=1}^{d} \tilde{h}_j(t) S_j^*(t).$$

To establish the theorem, it remains to be shown that $h_0^*(t) = \tilde{h}_0(t)$. From (7.3.35),

$$\tilde{h}_0(t) = e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right) - \sum_{j=1}^{d} \tilde{h}_j(t) S_j^*(t),$$

such that it suffices to show that

$$\mathcal{V}^{A^*}(t) - A^*(t) = e^{-\int_0^t (\gamma(u)r(u)+\delta(u))\,\mathrm{d}u}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right).$$

But this also immediately follows from Lemma 7.3.6 thus completing the proof. □

Proposition 7.3.7 allows us to draw the following conclusion. If a systemic investor assumes all payments, including taxes and expenses, of an original investor adopting the risk-minimizing investment strategy in the presence of taxes and expenses, and faces the problem of classic risk-minimization (in the absence of taxes and expenses), then the optimal strategy coincides with the investment strategy adopted by the original investor; in particular, additional risk reduction is impossible, and in this specific sense, tax- and expense-modified risk-minimization is consistent with classic risk-minimization.

7.4 Case study: classic multi-state life insurance payments

In Example 7.2.1 and Example 7.3.1, we considered an extension of the classic life insurance setting (confer with Hoem, 1969; Norberg, 1991; Christiansen, 2012) by allowing for investments in a bond market following a Vasicek term structure model with a deterministic tax rate and expenses depending on the current state of the insurance contract(s). This setup excluding taxes and expenses is similar to earlier examples from the literature on risk-minimization, see e.g. Møller (2001, Subsection 3.1).

In this section, we derive risk-minimizing investment strategies in the presence of taxes and expenses within the framework of Example 7.2.1 and Example 7.3.1 using the tools developed in Section 7.3. Throughout the exposition, we explain how to extend the results to general diffusion term structure models.

The models for the market, insurance payment process, and taxes and expenses were introduced in Example 7.2.1 and Example 7.3.1 and will not be repeated here. To keep the notation simple, we disregarded lump-sum payments in the examples and shall continue to do so here. An extension to more general payments is straightforward.

The section proceeds as follows. The risk-minimizing investment strategy in the presence of taxes and expenses is derived in Subsection 7.4.1, and valuation and computability with a view towards actuarial practice is discussed in Subsection 7.4.2.

7.4.1 Tax- and expense-modified risk-minimization

Based on the Markovianity of the short-rate r, define F and $F^{1-\gamma}$ by

$$F(t,r(t),s) = \mathrm{E}^Q\Big[e^{-\int_t^s r(u)\,\mathrm{d}u}\,\Big|\,r(t)\Big] = \mathrm{E}^Q\Big[e^{-\int_t^s r(u)\,\mathrm{d}u}\,\Big|\,\mathcal{F}(t)\Big],$$
$$F^{1-\gamma}(t,r(t),s) = \mathrm{E}^Q\Big[e^{-\int_t^s (1-\gamma)r(u)\,\mathrm{d}u}\,\Big|\,r(t)\Big] = \mathrm{E}^Q\Big[e^{-\int_t^s (1-\gamma)r(u)\,\mathrm{d}u}\,\Big|\,\mathcal{F}(t)\Big].$$

Note that $F(t,r(t),T) = S_1(t)$. Also note that

$$\mathrm{d}(1-\gamma)r(t) = \kappa\left((1-\gamma)\theta - (1-\gamma)r(t)\right)\mathrm{d}t + (1-\gamma)\sigma\,\mathrm{d}W(t),$$

so that $t \mapsto (1-\gamma)r(t)$ is another Ornstein-Uhlenbeck process. Using explicit results for the Vasicek term structure model, see e.g. Björk (2009, Proposition 24.3), we then find that

$$F_r^{1-\gamma}(t, r(t), s) = (1-\gamma) F_r(t, r(t), s) \frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)}, \tag{7.4.1}$$

where $F_r(t, r, s) = \frac{\partial}{\partial r} F(t, r, s)$ and similarly for $F_r^{1-\gamma}(t, r, s)$.

Define also so-called expense deflated transition probabilities $p^{-\delta}$ by

$$p_{ij}^{-\delta}(t, s) = \mathrm{E}^Q\Big[\mathbb{1}_{\{Z(s)=j\}} e^{\int_t^s \delta_{Z(u)}(u)\,\mathrm{d}u} \,\Big|\, Z(t) = i\Big].$$

Note that if $\delta \geq 0$, then $p_{ij}^{-\delta} \geq p_{ij}^{-0}$, where the latter are actually the ordinary transition probabilities, which satisfy Kolmogorov's backward and forward differential equations. In the general case, for $\delta \neq 0$, it follows from Appendix 7.A that the expense deflated transition probabilities satisfy systems of ordinary differential equations similar to Kolmogorov's backward and forward differential equations.

The tax- and expense-modified version $\tilde{A}^{\mathrm{b}}$ of the discounted insurance payment process is given by

$$\tilde{A}^{\mathrm{b}}(t) = A^{\mathrm{b}}(0) + \int_0^t e^{-\int_0^s ((1-\gamma)r(u) - \delta(u))\,\mathrm{d}u} \,\mathrm{d}A^{\mathrm{b}}(s),$$

confer with e.g. (7.3.30).

The following approach follows along the lines of Møller (2001, Section 3). Using the independence between Z and the financial market, it can be shown that the intrinsic value process associated with $\tilde{A}^{\mathrm{b}}$ can be written as

$$\begin{aligned} \mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) &= \mathrm{E}^Q\Big[\tilde{A}^{\mathrm{b}}(T) \,\Big|\, \mathcal{F}(t)\Big] & (7.4.2) \\ &= \tilde{A}^{\mathrm{b}}(t) + e^{-\int_0^t ((1-\gamma)r(u) - \delta(u))\,\mathrm{d}u} \int_t^T F^{1-\gamma}(t, r(t), s) Y_{Z(t)}^{-\delta}(t, s)\,\mathrm{d}s, & (7.4.3) \end{aligned}$$

where $Y_i^{-\delta}$ is given by

$$Y_i^{-\delta}(t, s) = \sum_{j \in \mathcal{J}} p_{ij}^{-\delta}(t, s) \left(b_j(s) + \sum_{k: k \neq j} \mu_{jk}(s) b_{jk}(s) \right).$$

Note that if $\delta \geq 0$, $b_j \geq 0$, and $b_{jk} \geq 0$, then $Y_i^{-\delta} \geq Y_i^{-0}$ since $p_{ij}^{-\delta} \geq p_{ij}^{-0}$; here Y_i^{-0} is the classic state-wise expected cash flow in the absence of expenses.

Since the derived expression for the intrinsic value process is comparable to that of Møller (2001, p. 426), we may proceed using the same techniques as in Møller (2001, pp. 442–444).

Define so-called state-wise prospective reserves $V_i^{1-\gamma,\delta}$ by

$$V_i^{1-\gamma,\delta}(t) = \int_t^T F^{1-\gamma}(t, r(t), s) Y_i^{-\delta}(t, s) \, \mathrm{d}s. \tag{7.4.4}$$

We are now ready to state the relevant Galtchouk-Kunita-Watanabe decomposition:

Lemma 7.4.1. *The Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\tilde{A}^{\mathrm{b}}}$ is given by*

$$\mathcal{V}^{\tilde{A}^b}(t) = \mathcal{V}^{\tilde{A}^b}(0) + \int_0^t \xi_{Z(s-)}(s) \, \mathrm{d}S_1^*(s) + \sum_{j \in \mathcal{J}} \sum_{k:k \neq j} \int_0^t v_{jk}(s) \, \mathrm{d}M_{jk}(s).$$

where

$$\xi_i(t) = (1-\gamma) e^{\int_0^t (\gamma r(u) + \delta(u)) \, \mathrm{d}u} \int_t^T \frac{F_r(t, r(t), s)}{F_r(t, r(t), T)} \frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)} Y_i^{-\delta}(t, s) \, \mathrm{d}s, \tag{7.4.5}$$

$$v_{jk}(t) = e^{-\int_0^t ((1-\gamma) r(u) - \delta(u)) \, \mathrm{d}u} \left(b_{jk}(t) + V_k^{1-\gamma,\delta}(t) - V_j^{1-\gamma,\delta}(t) \right). \tag{7.4.6}$$

Proof. The proof mirrors the proof of Møller (2001, Lemma 3.2), although under relaxed regularity conditions. We therefore only sketch the essential steps with a focus on the complications that arrive due to the inclusion of taxes and expenses.

First, one takes a closer look at the dynamics of

$$e^{-\int_0^t ((1-\gamma) r(u) - \delta(u)) \, \mathrm{d}u} V_i^{1-\gamma,\delta}(t).$$

Using a system of ordinary differential equations similar to Kolmogorov's backward differential equations, see Appendix 7.A, and then proceeding along the lines of Møller (2001, pp. 443–444), it is then possible to show that

$$\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) = \mathcal{V}^{\tilde{A}^{\mathrm{b}}}(0) + \int_0^t \tilde{\xi}_{Z(s-)}(s) \, \mathrm{d}W(s) + \sum_{j \in \mathcal{J}} \sum_{k:k \neq j} \int_0^t v_{jk}(s) \, \mathrm{d}M_{jk}(s),$$

where v_{jk} is given by (7.4.6) and

$$\tilde{\xi}_i(t) = e^{-\int_0^t ((1-\gamma) r(u) - \delta(u)) \, \mathrm{d}u} \sigma \int_t^T F_r^{1-\gamma}(t, r(t), s) Y_i^{-\delta}(t, s) \, \mathrm{d}s. \tag{7.4.7}$$

Using the Vasicek term structures, we next find that

$$\begin{aligned}
\tilde{\xi}_i(t) &= e^{-\int_0^t ((1-\gamma) r(u) - \delta(u)) \, \mathrm{d}u} (1-\gamma) \sigma \int_t^T F_r(t, r(t), s) \frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)} Y_i^{-\delta}(t, s) \, \mathrm{d}s \\
&= e^{\int_0^t (\gamma r(u) + \delta(u)) \, \mathrm{d}u} (1-\gamma) \int_t^T \frac{F_r(t, r(t), s)}{F_r(t, r(t), T)} \frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)} Y_i^{-\delta}(t, s) \, \mathrm{d}s \\
&\quad e^{-\int_0^t r(s) \, \mathrm{d}s} F_r(t, r(t), T) \, \sigma,
\end{aligned}$$

see also (7.4.1). In other words,

$$\tilde{\xi}_i(t) = \xi_i(t)e^{-\int_0^t r(s)\,\mathrm{d}s} F_r(t, r(t), T)\,\sigma$$

with ξ_i defined by (7.4.5). Now recall that

$$\mathrm{d}S_1^*(t) = e^{-\int_0^t r(s)\,\mathrm{d}s} F_r(t, r(t), T)\,\sigma\,\mathrm{d}W(t),$$

from which it follows that

$$\tilde{\xi}_i(t)\,\mathrm{d}W(t) = \xi_i(t)\,\mathrm{d}S_1^*(t),$$

completing the sketch of proof. □

Remark 7.4.2. For a general diffusion short rate model, the proof technique of Lemma 7.4.1 still applies and similar results as in the Vasicek term structure model remain obtainable. Assume the short rate satisfies the stochastic differential equation

$$\mathrm{d}r(t) = \alpha(t, r(t))\,\mathrm{d}t + \sigma(t, r(t))\,\mathrm{d}W(t),$$

with W still a standard Brownian motion under Q and where α and σ are functions satisfying certain Lipschitz conditions. Imposing suitable additional regularity conditions on the short rate model (equivalently, the term structure model), one finds

$$\begin{aligned}
\tilde{\xi}_i(t) &= e^{-\int_0^t((1-\gamma)r(u)-\delta(u))\,\mathrm{d}u}\sigma(t, r(t)) \int_t^T F_r^{1-\gamma}(t, r(t), s) Y_i^{-\delta}(t, s)\,\mathrm{d}s \\
&= e^{\int_0^t(\gamma r(u)+\delta(u))\,\mathrm{d}u} \int_t^T \frac{F_r^{1-\gamma}(t, r(t), s)}{F_r(t, r(t), T)} Y_i^{-\delta}(t, s)\,\mathrm{d}s \\
&\quad e^{-\int_0^t r(s)\,\mathrm{d}s} F_r(t, r(t), T)\,\sigma(t, r(t)),
\end{aligned}$$

by following along the lines of the sketch of proof of Lemma 7.4.1. This results in the following Galtchouk-Kunita-Watanabe decomposition:

$$\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) = \mathcal{V}^{\tilde{A}^{\mathrm{b}}}(0) + \int_0^t \mathbb{1}_{\{Z(s-)=i\}}\xi_i(s)\,\mathrm{d}S_1^*(s) + \sum_{j\in\mathcal{J}}\sum_{k:k\neq j}\int_0^t v_{jk}(s)\,\mathrm{d}M_{jk}(s).$$

where now

$$\begin{aligned}
\xi_i(t) &= e^{\int_0^t(\gamma r(u)+\delta(u))\,\mathrm{d}u} \int_t^T \frac{F_r^{1-\gamma}(t, r(t), s)}{F_r(t, r(t), T)} Y_i^{-\delta}(t, s)\,\mathrm{d}s, \\
v_{jk}(t) &= e^{-\int_0^t((1-\gamma)r(u)-\delta(u))\,\mathrm{d}u}\left(b_{jk}(t) + V_k^{1-\gamma,\delta}(t) - V_j^{1-\gamma,\delta}(t)\right).
\end{aligned}$$ ▽

As we have identified the Galtchouk-Kunita-Watanabe decomposition of $\mathcal{V}^{\tilde{A}^{\mathrm{b}}}$, we are now ready to apply the results on tax- and expense-modified risk-minimization to obtain the main result of this section:

Theorem 7.4.3. *The unique risk-minimizing investment strategy $\tilde{h}$ in the setting of Section 7.4 is given as follows:*

$$\tilde{h}_1(t) = \int_t^T \frac{F_r(t,r(t),s)}{F_r(t,r(t),T)} \frac{F^{1-\gamma}(t,r(t),s)}{F(t,r(t),s)} Y^{-\delta}_{Z(t-)}(t,s)\,\mathrm{d}s,$$
$$\tilde{h}_0(t) = S_0^{-1}(t)\left(V^{1-\gamma,\delta}_{Z(t)}(t) - \tilde{h}_1(t)S_1(t)\right).$$

The associated value process is

$$V(\tilde{h},t) = V^{1-\gamma,\delta}_{Z(t)}(t).$$

Proof. The first statement follows immediately by combining Lemma 7.4.1 with Theorem 7.3.5 and the observation

$$\begin{aligned} e^{-\int_0^t (\gamma r(u)+\delta(u))\,\mathrm{d}u}\left(\mathcal{V}^{\tilde{A}^{\mathrm{b}}}(t) - \tilde{A}^{\mathrm{b}}(t)\right) &= S_0^{-1}(t)\int_t^T F^{1-\gamma}(t,r(t),s)Y^{-\delta}_{Z(t)}(t,s)\,\mathrm{d}s \\ &= S_0^{-1}(t)V^{1-\gamma,\delta}_{Z(t)}(t), \end{aligned}$$

confer with (7.4.3). The last statement follows by direct calculations from the first statement and (7.2.3). □

Remark 7.4.4. In Remark 7.4.2 we discussed extensions of the Galtchouk-Kunita-Watanabe decomposition of Lemma 7.4.1 to general diffusion short rate models. Based on this discussion, we can extend the conclusions of Theorem 7.4.3 to the general framework of Remark 7.4.2 in the following manner.

Assume the short rate satisfies the stochastic differential equation

$$\mathrm{d}r(t) = \alpha(t,r(t))\,\mathrm{d}t + \sigma(t,r(t))\,\mathrm{d}W(t),$$

with W still a standard Brownian motion under Q and where α and σ are functions satisfying certain Lipschitz conditions. Imposing suitable additional regularity conditions, the unique risk-minimizing investment strategy $\tilde{h}$ is given by

$$\tilde{h}_1(t) = \int_t^T \frac{1}{1-\gamma} \frac{F_r^{1-\gamma}(t,r(t),s)}{F_r(t,r(t),T)} Y^{-\delta}_{Z(t-)}(t,s)\,\mathrm{d}s, \tag{7.4.8}$$
$$\tilde{h}_0(t) = S_0^{-1}(t)\left(V^{1-\gamma,\delta}_{Z(t)}(t) - \tilde{h}_1(t)S_1(t)\right).$$

The associated value process is still

$$V(\tilde{h},t) = V^{1-\gamma,\delta}_{Z(t)}(t).$$ ▽

7.4.2 Discussion

The value process $V(\tilde{h},t) = V^{1-\gamma,\delta}_{Z(t)}(t)$ associated with the risk-minimizing investment strategy is affected by the introduction of taxes and expenses. In the following, we focus on the benefit part of the paymend process A^{b} by requiring $b_j \geq 0$ and $b_{jk} \geq 0$. From the discussion leading up to (7.4.4), we can conclude that the value process is increasing in expenses, δ. With regards to taxes, γ, the effect is not as clear. First note that the after-tax zero coupon bond prices take the form

$$F^{1-\gamma}(t, r(t), s) = \mathrm{E}^Q\left[e^{-\int_t^s (1-\gamma) r(u)\,\mathrm{d}u} \,\middle|\, r(t)\right],$$

and if $r > 0$, it is evident that $F^{1-\gamma}(t, r(t), s)$ is increasing in γ, and in that case, the value process increases as the level of taxation increases. In the general case, where the interest rate is allowed to be negative, the effect of taxes on the value process depends on whether r, i.e. the return on the savings account, is mostly positive or negative. Consequently, a general (model independent) statement seems out of reach.

The unique risk-minimizing investment strategy $\tilde{h}$ given by Theorem 7.4.3 is a modification of the classic strategy without taxes and expenses. The quantity $Y^{-\delta}_{Z(t-)}(t, s)$ is the expected (think: diversified with respect to future unsystematic insurance risk) rate of payments at time s given the present state of the insurance contract(s) at time t while taking future state-wise expenses into account; it can be interpreted as an expected expense-modified cash flow.

The integrand in the expression of $\tilde{h}_1$ from Theorem 7.4.3,

$$s \mapsto \frac{F_r(t, r(t), s)}{F_r(t, r(t), T)} \frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)} Y^{-\delta}_{Z(t-)}(t, s), \tag{7.4.9}$$

is the rate of investment into the bond which is dictated by the risk-minimizing strategy to cover the future expected payments. Thus, in regards to investment into the risky asset, taxes are taken into account by scaling the investment by a factor of

$$\frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)},$$

confer also with the discussion in Buchardt and Møller (2018, Section 4, in particular Subsection 4.3.2).

The product structure of (7.4.9) w.r.t. taxes and expenses is a direct consequence of the independence between market and insurance risks and the fact that tax rate does not depend on the history of the insurance contract(s) nor the market while the expense rate only depends on the history of the insurance contract(s).

To explicitly compute the risk-minimizing investment strategy and the associated value process, one needs to calculate F, F_r, and $F^{1-\gamma}$ as well as the expense deflated transition probabilities $p^{-\delta}_{ij}$. For the Vasicek term structure model, where in particular $t \mapsto (1-\gamma)r$ is another Ornstein-Uhlenbeck process, the former quantities have closed-form expressions and are therefore easily calculated. Because the expense deflated transition probabilities $p^{-\delta}_{ij}$ can be found by solving a system of ordinary differential equations similar to Kolmogorov's forward differential equations, see Appendix 7.A, this establishes a simple scheme for the computation of the risk-minimizing investment strategy and the associated value process.

In Remark 7.4.4 we elaborated on how to extend Theorem 7.4.3 to general diffusion short rate models. If the model is affine, the relevant quantities needed for computation of the risk-minimizing investment strategy and the associated value process, i.e. F, $F_r^{1-\gamma}$, and $F^{1-\gamma}$, can be calculated by solving systems of ordinary differential equations, see Duffie, Pan, and Singleton (2000) and Buchardt (2016). The model of Section 7.4 can therefore easily be implemented in practice; and furthermore, the extension to continuous affine term structure models is relatively straightforward.

In general diffusion short rate models, the scaling factor

$$\frac{F^{1-\gamma}(t, r(t), s)}{F(t, r(t), s)}$$

of the Vasicek model is replaced by

$$\frac{1}{1-\gamma} \frac{F_r^{1-\gamma}(t, r(t), s)}{F_r(t, r(t), s)} = \frac{F_{(1-\gamma)r}^{1-\gamma}(t, r(t), s)}{F_r(t, r(t), s)},$$

compare the results of Theorem 7.4.3 to (7.4.8). Since $F^{1-\gamma}$ is the price process of an artificial zero coupon bond based on a modified short rate $(1-\gamma)r$, this is the ratio of two (model dependent) interest rate sensitivities. We conclude that both qualitatively and quantitatively, how taxes affect the amount invested in the risky asset (the zero coupon bond) is non-trivial and model dependent.

For the Cox-Ingersoll-Ross model, where the interest rate is ensured positive, it is possible to show that

$$\frac{F_{(1-\gamma)r}^{1-\gamma}(t, r(t), s)}{F_r(t, r(t), s)} \geq 1.$$

The details of this derivation are omitted. For additional theoretical considerations as well as numerical results for the Vasicek and Cox-Ingersoll-Ross models, we refer to Buchardt and Møller (2018, Sections 4-5). While further investigations lie outside the scope of this paper, a general investigation could focus on affine models, since the tax modification retains affinity.

Acknowledgments and declarations of interest

Christian Furrer's research is partly funded by the Innovation Fund Denmark under File No. 7038-00007B. The authors declare no competing interests. We thank the editor and reviewer for comments that helped improve the paper.

7.A Deflated transition probabilities

Let $(\Omega, \mathcal{F}, P)$ be a background probability space, and let Z be a Markovian jump process with values in a finite set $\mathcal{J} = \{0, 1, \dots, n-1\}$. Let N be the multivariate counting process associated with Z. The transition probabilities of Z are given by $n \times n$-matrices $p(t,s)$ for $0 \leq t \leq s < \infty$, where

$$p_{ij}(t,s) = P[\, Z(s) = j \mid Z(t) = i],$$

and the transition probabilities satisfy the Chapman-Kolmogorov equation.

We assume the existence of continuous transition intensities μ_{jk}, when each counting process N_{jk} has intensity process λ_{jk} given by

$$\lambda_{jk}(t) = \mathbb{1}_{\{Z(t-)=j\}}\mu_{jk}(t).$$

We can then take regular versions of the conditional distributions for which the transition probabilities p satisfy

$$\mu(t) = \lim_{h \searrow 0} \frac{p(t, t+h) - p(t,t)}{h},$$

where μ are $n \times n$-matrices with diagonal elements

$$\mu_{jj} = -\sum_{k:k\neq j} \mu_{jk}.$$

Furthermore, the transition probabilities satisfy Kolmogorov's backward and forward differential equations.

Let δ be $n \times 1$-dimensional with deterministic and continuous elements $t \mapsto \delta_i(t)$. Quantities of interest are (corresponding regular versions) of

$$p_{ij}^{\delta}(t,s) = \mathrm{E}\left[\mathbb{1}_{\{Z(s)=j\}} e^{-\int_t^s \delta_{Z(u)}(u)\,\mathrm{d}u} \,\middle|\, Z(t) = i \right],$$

which we term *δ-deflated transition probabilities*. When $\delta \equiv 0_{n\times 1}$, we see that these quantities are in fact the transition probabilities. When $\delta \equiv 1_{n\times 1} f$ for some deterministic and continuous function $t \mapsto f(t)$, we see that

$$p_{ij}^{\delta}(t,s) = e^{-\int_t^s f(u)\,\mathrm{d}u} p_{ij}(t,s).$$

The δ-deflated transition probabilities satisfy systems of ordinary differential equations similar to Kolmogorov's backward and forward differential equations:

Lemma 7.A.1. *The δ-deflated transition probabilities satisfy the forward ordinary differential equation system*

$$\frac{\partial}{\partial s} p^{\delta}(t,s) = p^{\delta}(t,s)\,[\mu - \mathrm{diag}(\delta)](s),$$

and the backward ordinary differential equation system

$$\frac{\partial}{\partial t}p^{\delta}(t,s) = -\left[\mu - \mathrm{diag}(\delta)\right](t)p^{\delta}(t,s),$$

with boundary conditions $p^{\delta}(t,t) = \mathrm{diag}(1_{n\times 1})$.

Proof. The boundary conditions are evident. We first prove the forward differential equations. Define the $1\times n$-dimensional indicator process I by

$$I_i(t) = \mathbb{1}_{\{Z(t)=i\}}.$$

For fixed $t_0 \geq 0$, define also the $1\times n$-dimensional process X by

$$X(t) = I(t)e^{-\int_{t_0}^{t}\delta_{Z(u)}(u)\,\mathrm{d}u}.$$

View N as $n\times n$-matrices with diagonal elements

$$N_{jj} = -\sum_{k:k\neq j} N_{jk}$$

In similar fashion, view λ as $n\times n$-matrices with diagonal elements

$$\lambda_{jj} = -\sum_{k:k\neq j}\lambda_{jk},$$

such that $\lambda_{jj}(t) = I_j(t-)\mu_{jj}(t)$.

Recalling that $\mathrm{d}I_i = \sum_{j\neq i}(\mathrm{d}N_{ji} - \mathrm{d}N_{ij}) = \sum_j \mathrm{d}N_{ji}$, we see that

$$\mathrm{d}I = 1_{1\times n}\,\mathrm{d}N.$$

Because the compensated jump processes

$$t \mapsto N_{ij}(t) - \int_0^t \lambda_{ij}(s)\,\mathrm{d}s$$

are martingales, we find that

$$\begin{aligned}\mathrm{d}I(t) &= \mathrm{d}M(t) + 1_{1\times n}\lambda(t)\,\mathrm{d}t\\ &= \mathrm{d}M(t) + I(t-)\mu(t)\,\mathrm{d}t\\ &= \mathrm{d}M(t) + I(t)\mu(t)\,\mathrm{d}t,\end{aligned}\tag{7.A.1}$$

where M is a $1\times n$-dimensional martingale given by

$$\mathrm{d}M(t) = 1_{1\times n}\left(\mathrm{d}N(t) - \lambda(t)\,\mathrm{d}t\right).$$

Integration by parts now yields

$$\begin{aligned}\mathrm{d}X(t) &= \left(\mathrm{d}I(t)\right)e^{-\int_{t_0}^{t}\delta_{Z(u)}(u)\,\mathrm{d}u} - I(t)\delta_{Z(t)}(t)e^{-\int_{t_0}^{t}\delta_{Z(u)}(u)\,\mathrm{d}u}\,\mathrm{d}t\\ &= X(t)\mu(t)\mathrm{d}t - X(t)\delta_{Z(t)}(t)\,\mathrm{d}t + e^{-\int_{t_0}^{t}\delta_{Z(u)}(u)\,\mathrm{d}u}\,\mathrm{d}M(t).\end{aligned}$$

By definition of X,

$$\mathrm{E}\left[X_j(t)\middle|\, Z(t_0)=i\right] = p^{\delta}_{ij}(t_0,t),$$
$$\mathrm{E}\left[\delta_{Z(t)}(t)X_j(t)\middle|\, Z(t_0)=i\right] = \delta_j(t)p^{\delta}_{ij}(t_0,t).$$

The latter corresponds to the (i,j)'th element of the matrix product of $p^{\delta}(t_0,t)$ and the diagonal matrix with diagonal $\delta(t)$. Collecting all terms, it then follows from Fubini's theorem and the martingale properties of M that

$$p^{\delta}(t_0,t) = p^{\delta}(t_0,t_0) + \int_{t_0}^{t} p^{\delta}(t_0,s)\,[\mu - \mathrm{diag}(\delta)](s)\,\mathrm{d}s.$$

The forward differential equations now follow by differentiation w.r.t. t.

We now turn our attention to the backward differential equations. For fixed $s \geq 0$ define the $1 \times n$-dimensional martingale Y by

$$\begin{aligned} Y(t) &= \mathrm{E}\left[I(s)e^{-\int_0^s \delta_{Z(u)}(u)\,\mathrm{d}u}\middle|\, \mathcal{F}(t)\right] \\ &= I(t)e^{-\int_0^t \delta_{Z(u)}(u)\,\mathrm{d}u} p^{\delta}(t,s), \end{aligned}$$

where $0 \leq t \leq s$.

Integration by parts now yields

$$\begin{aligned} e^{\int_0^t \delta_{Z(u)}(u)\,\mathrm{d}u}\,\mathrm{d}Y(t) =& \,\mathrm{d}\big(I(t)p^{\delta}(t,s)\big) - I(t)\delta_{Z(t)}(t)p^{\delta}(t,s)\,\mathrm{d}t \\ =& \, I(t)p^{\delta}(\mathrm{d}t,s) + I(t)\mu(t)p^{\delta}(t,s)\,\mathrm{d}t - I(t)\delta_{Z(t)}(t)p^{\delta}(t,s)\,\mathrm{d}t \\ & + \mathrm{d}M(t)p^{\delta}(t,s), \end{aligned}$$

where we have used (7.A.1). Because Y and M are $1 \times n$-dimensional martingales, we find using martingale representation theory that $p^{\delta}(t,s)$ is differentiable in t and that

$$I(t)\frac{\partial}{\partial t}p^{\delta}(t,s) = -\big(I(t)\mu(t)p^{\delta}(t,s) - I(t)\delta_{Z(t)}(t)p^{\delta}(t,s)\big).$$

The backward differential equations can now be established by taking a closer look at this expression on each event $\{Z(t)=i\}$ for varying i. For example, on $\{Z(t)=i\}$,

$$\big[I(t)\delta_{Z(t)}(t)p^{\delta}(t,s)\big]_j = \delta_i(t)p^{\delta}_{ij}(t,s),$$

which corresponds to the (i,j)'th element of the matrix product between the diagonal matrix with diagonal $\delta(t)$ and the matrix $p^{\delta}(t,s)$. By additional observations of the same kind, the backward differential equations follow. This completes the proof. □

Bibliography

Aalen, O.O., Ø. Borgan, and H.K. Gjessing (2008). *Survival and Event History Analysis*. Statistics for Biology and Health. Springer-Verlag New York. DOI: 10.1007/978-0-387-68560-1.

Adékambi, F. and M.C. Christiansen (2017). Integral and differential equations for the moments of multistate models in health insurance. *Scandinavian Actuarial Journal* **2017**, pp. 29–50. DOI: 10.1080/03461238.2015.1058854.

Ahmad, J., K. Buchardt, and C. Furrer (2020). *Computation of bonus in multi-state life insurance*. arXiv: 2007.04051.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemporary Mathematics* **80**, pp. 19–60. DOI: 10.1090/conm/080/999006.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York. DOI: 10.1007/978-1-4612-4348-9.

Asmussen, S. and M. Steffensen (2020). *Risk and Insurance*. Probability Theory and Stochastic Modelling. Springer, Cham. DOI: 10.1007/978-3-030-35176-2.

Battauz, A. (2003). Quadratic hedging for asset derivatives with discrete stochastic dividends. *Insurance: Mathematics and Economics* **32**(2), pp. 229–243. DOI: 10.1016/S0167-6687(02)00212-3.

Bauer, D., F.E. Benth, and R. Kiesel (2012). Modeling the Forward Surface of Mortality. *SIAM Journal on Financial Mathematics* **3**, 639–666. DOI: 10.1137/100818261.

Biagini, F., A. Groll, and J. Widenmann (2016). Risk Minimization for Insurance Products via F-Doubly Stochastic Markov Chains. *Risks* **4**(3), p. 23. DOI: 10.3390/risks4030023.

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. 3rd ed. Oxford Finance Series. Oxford University Press, Oxford. DOI: 10.1093/0199271267.001.0001.

Bladt, M., S. Asmussen, and M. Steffensen (2020). Matrix representations of life insurance payments. *European Actuarial Journal*. DOI: 10.1007/s13385-019-00222-0.

Brennan, M.J. and E.S. Schwartz (1979a). Alternative Investment Strategies for the Issuers of Equity Linked Life Insurance Policies with an Asset Value Guarantee. *The Journal of Business* **52**, pp. 63–93. DOI: 10.1086/296034.

Brennan, M.J. and E.S. Schwartz (1979b). *Pricing and Investment Strategies for Guaranteed Equity-Linked Life Insurance*. S.S. Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania, Philadelphia.

Bruhn, K. and A.S. Lollike (2020). *Retrospective Reserves and Bonus*. Preprint. URL: http://dx.doi.org/10.2139/ssrn.3560910.

Buchardt, K. (2014). Dependent interest and transition rates in life insurance. *Insurance: Mathematics and Economics* **55**, pp. 167–179. DOI: 10.1016/j.insmatheco.2014.01.004.

Buchardt, K. (2016). Continuous affine processes: transformations, Markov chains and life insurance. *Advances in Applied Probability* **48**(2), pp. 423–442. DOI: 10.1017/apr.2016.8.

Buchardt, K. (2017). Kolmogorov's forward PIDE and forward transition rates in life insurance. *Scandinavian Actuarial Journal* **2017**(5), pp. 377–394. DOI: 10.1080/03461238.2016.1160255.

Buchardt, K., C. Furrer, and T. Møller (2020). Tax- and expense-modified risk-minimization for insurance payment processes. *Scandinavian Actuarial Journal*, to appear. DOI: 10.1080/03461238.2020.1790413.

Buchardt, K., C. Furrer, and M. Steffensen (2019). Forward transition rates. *Finance and Stochastics* **23**(4), 975–999. DOI: 10.1007/s00780-012-0182-3.

Buchardt, K. and T. Møller (2015). Life Insurance Cash Flows with Policyholder Behavior. *Risks* **3**(3), 290–317. DOI: 10.3390/risks3030290.

Buchardt, K. and T. Møller (2018). Hedging and Cash Flows in the Presence of Taxes and Expenses in Life and Pension Insurance. *Risks* **6**(3). DOI: 10.3390/risks6030068.

Buchardt, K., T. Møller, and K.B. Schmidt (2015). Cash flows and policyholder behaviour in the semi-Markov life insurance setup. *Scandinavian Actuarial Journal* **2015**(8), pp. 660–688. DOI: 10.1080/03461238.2013.879919.

Bühlmann, H. (1989). Actuaries of the Third Kind? *ASTIN Bulletin* **19**(S1), pp. 5–6. DOI: 10.1017/S0515036100008576.

Bühlmann, H. and E. Straub (1970). Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* **70**, pp. 111–133.

Christiansen, M.C. (2006). *A joint analysis of financial and biometrical risks in life insurance*. PhD thesis. University of Rostock.

Christiansen, M.C. (2012). Multistate models in health insurance. *Advances in Statistical Analysis* **96**, pp. 155–186. DOI: `10.1007/s10182-012-0189-2`.

Christiansen, M.C. (2020). *A martingale concept for non-monotone information in a jump process framework*. arXiv: `1811.00952`.

Christiansen, M.C. and B. Djehiche (2020). Nonlinear reserving and multiple contract modifications in life insurance. *Insurance: Mathematics and Economics* **93**, pp. 187–195. DOI: `10.1016/j.insmatheco.2020.05.004`.

Christiansen, M.C. and C. Furrer (2020). *Dynamics of state-wise prospective reserves in the presence of non-monotone information*. arXiv: `2003.02173`.

Christiansen, M.C. and A. Niemeyer (2015). On the forward rate concept in multi-state life insurance. *Finance and Stochastics* **19**(2), pp. 295–327. DOI: `10.1007/s00780-014-0244-9`.

Christiansen, M.C. and E. Schinzinger (2016). A credibility approach for combining likelihood of generalized linear models. *ASTIN Bulletin* **46**(3), pp. 531–569. DOI: `10.1017/asb.2016.11`.

Cohen, S.N. and R.J. Elliott (2012). Existence, uniqueness and comparisons for BSDEs in general spaces. *The Annals of Probability* **40**(5), pp. 2264–2297. DOI: `10.1214/11-AOP679`.

Dahl, M. (2004). Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics* **35**, pp. 113–136. DOI: `10.1016/j.insmatheco.2004.05.003`.

Dahl, M. and T. Møller (2006). Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: Mathematics and Economics* **39**(2), pp. 193–217. DOI: `10.1016/j.insmatheco.2006.02.007`.

Deprez, P., P.V. Shevchenko, and M.V. Wüthrich (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal* **7**(2), pp. 337–352. DOI: `10.1007/s13385-017-0152-4`.

Djehiche, B. and B. Löfdahl (2016). Nonlinear reserving in life insurance: Aggregation and mean-field approximation. *Insurance: Mathematics and Economics* **69**(3), pp. 1–13. DOI: `10.1016/j.insmatheco.2016.04.002`.

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* **68**(6), 1343–1376. DOI: `10.3386/w7105`.

European Insurance and Occupational Pensions Authority (2009). *Solvency II Directive (Directive 2009/138/EC)*. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009L0138-20140523&from=EN`.

European Insurance and Occupational Pensions Authority (2015). *Commission Delegated Regulation (EU) 2015/35 of 10 October 2014 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2015:012:FULL&from=EN`.

Falden, D. and A.K. Nyegaard (2020). *Retrospective Reserves and Bonus with Policyholder Behavior*. Preprint. URL: `https://dx.doi.org/10.2139/ssrn.3575544`.

Föllmer, H. and D. Sondermann (1986). "Hedging of Non-Redundant Contingent Claims". In: *Contributions to Mathematical Economics*. Ed. by W. Hildenbrand and A. Mas-Colell. Elsevier Science, North-Holland, pp. 205–223. DOI: `10.13140/RG.2.1.3298.8322`.

Furrer, C. (2019). Experience rating in the classic Markov chain life insurance setting. *European Actuarial Journal* **9**, pp. 31–58. DOI: `https://doi.org/10.1007/s13385-019-00190-5`.

Gad, K.S.T. and J.W. Nielsen (2016). Reserves and cash flows under stochastic retirement. *Scandinavian Actuarial Journal* **2016**(10), pp. 876–904. DOI: `10.1080/03461238.2015.1028432`.

Gerds, T.A. and M. Schumacher (2006). Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biometrical Journal* **48**(6), pp. 1029–1040. DOI: `10.1002/bimj.200610301`.

Ghitany, M.E., D. Karlis, D.K. Mutairi, and F.A. Al-Awadhi (2012). An EM algorithm for multivariate mixed Poisson regression models and its application. *Applied Mathematical Sciences* **6**(137), pp. 6843–6856.

Gill, R.D. and S. Johansen (1990). A Survey of Product-Integration with a View Toward Application in Survival Analysis. *Annals of Statistics* **18**(4), pp. 1501–1555. DOI: `10.1214/aos/1176347865`.

Gschlössl, S., P. Schoenmaekers, and M. Denuit (2011). Risk classification in life insurance: methodology and case study. *European Actuarial Journal* **1**, pp. 23–41. DOI: `10.1007/s13385-011-0028-y`.

Guasoni, P. (2002). Risk minimization under transaction costs. *Finance and Stochastics* **6**, pp. 91–113. DOI: `10.1007/s780-002-8402-0`.

Haastrup, S. (2000). Comparison of Some Bayesian Analyses of Heterogeneity in Group Life Insurance. *Scandinavian Actuarial Journal* **2000**, pp. 2–16. DOI: 10.1080/034612300750066692.

Heath, D., R. Jarrow, and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates. *Econometrica* **60**, pp. 77–106. DOI: 10.2307/2951677.

Helwich, M. (2008). *Durational effects and non-smooth semi-Markov models in life insurance*. PhD thesis. University of Rostock.

Henriksen, L.F.B. (2014). *Aspects of valuation and optimization in life insurance*. PhD thesis. University of Copenhagen. URL: http://web.math.ku.dk/noter/filer/phd14lfbh.pdf.

Henriksen, L.F.B, J.W. Nielsen, M. Steffensen, and C. Svensson (2014). Markov chain modeling of policyholder behavior in life insurance and pension. *European Actuarial Journal* **4**, pp. 1–29. DOI: 10.1007/s13385-014-0091-2.

Hoem, J.M. (1969). Markov Chain Models in Life Insurance. *Blätter der DGVFM* **9**, pp. 91–107. DOI: 10.1007/BF02810082.

Hoem, J.M. (1972). Inhomogeneous Semi-Markov Processes, Select Actuarial Tables, and Duration-Dependence in Demography. In: *Population Dynamics*. Ed. by T.N.E. Greville. Academic Press, pp. 251–296. DOI: 10.1016/B978-1-4832-2868-6.50013-8.

Hoem, J.M. and O.O. Aalen (1978). Actuarial values of payment streams. *Scandinavian Actuarial Journal* **1978**, pp. 38–47. DOI: 10.1080/03461238.1978.10414317.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. Springer, New York. DOI: 10.1007/978-1-4612-1304-8.

International Accounting Standards Board (2017). *IFRS 17 Insurance Contracts*. https://www.idc.ac.il/he/specialprograms/accounting/fvf/Documents/IFRS17/WEBSITE155.pdf.

Jacobsen, M. (1982). *Statistical Analysis of Counting Processes*. Lecture Notes in Statistics. Springer, New York. DOI: 10.1007/978-1-4684-6275-3.

Jacobsen, M. (2006). *Point process theory and applications: Marked point and piecewise deterministic processes*. Probability and its Applications. Birkhäuser, Boston. DOI: 10.1007/0-8176-4463-6.

Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **31**, pp. 235–253. DOI: 10.1007/BF00536010.

Jacod, J. and A.N. Shiryaev (2003). *Limit Theorems for Stochastic Processes.* 2nd ed. Grundlehren der mathematischen Wissenschaften. Springe, Berlin, Heidelberg. DOI: 10.1007/978-3-662-05265-5.

Janssen, J. (1966). Application des processus semi-markoviens à un problème d'invaliditè. *Bulletin de l'Association Royale des Actuaires Belges* **63**, pp. 35–52.

Jarner, S.F. and T. Møller (2015). A partial internal model for longevity risk. *Scandinavian Actuarial Journal* **2015**(4), pp. 352–382. DOI: 10.1080/03461238.2013.836561.

Jensen, N. and K. Schomacker (2015). A Two-Account Life Insurance Model for Scenario-Based Valuation Including Event Risk. *Risks* **3**(2), pp. 183–218. DOI: 10.3390/risks3020183.

Jensen, N.R. (2016). Scenario-based life insurance prognoses in a multi-state Markov model. *European Actuarial Journal* **6**, pp. 307–330. DOI: 10.1007/s13385-016-0137-8.

Kallenberg, O. (1997). *Foundations of Modern Probability.* 1st ed. Probability and its Applications. Springer, New York. DOI: 10.1007/b98838.

Klugman, S., T. Rhodes, M. Purushotham, S. Gill, and MIB Solutions (2009). *Credibility Theory Practices.* Society of Actuaries.

Kraft, H. and M. Steffensen (2007). Bankruptcy, Counterparty Risk, and Contagion. *Review of Finance* **11**, pp. 209–252. DOI: 10.1093/rof/rfm002.

Lamberton, D., H. Pham, and M. Schweizer (1998). Local Risk-Minimization under Transaction Costs. *Mathematics of Operations Research* **23**(3), pp. 585–612. DOI: 10.1287/moor.23.3.585.

Last, G. and A. Brandt (1995). *Marked Point Processes on the Real Line.* Probability and Its Applications. Springer, New York.

Lee, Y. and J.A. Nelder (1996). Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(4), pp. 619–678. DOI: 10.1111/j.2517-6161.1996.tb02105.x.

Lewis, P.A. and G.S. Shedler (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* **26**(3), pp. 403–413. DOI: 10.1002/nav.3800260304.

Milbrodt, H. and A. Stracke (1997). Markov models and Thiele's integral equations for the prospective reserve. *Insurance: Mathematics and Economics* **19**(3), pp. 187–235. DOI: 10.1016/S0167-6687(97)00020-6.

Milevsky, M. and D. Promislow (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics* **29**(3), pp. 299–318. DOI: 10.1016/S0167-6687(01)00093-2.

Miltersen, K. and S. Persson (2005). *Is mortality dead? Stochastic forward force of mortality rate determined by no arbitrage*. Working paper, Norwegian School of Economics and Business Administration. URL: http://www.mathematik.uni-ulm.de/carfi/vortraege/downloads/DeadMort.pdf.

Møller, C.M. (1993). A stochastic version of Thiele's differential equation. *Scandinavian Actuarial Journal* **1993**, pp. 1–16. DOI: 10.1080/03461238.1993.10413910.

Møller, T. (1998). Risk-minimizing hedging strategies for unit-linked life insurance contracts. *ASTIN Bulletin* **28**, pp. 17–47. DOI: 10.2143/AST.28.1.519077.

Møller, T. (2001). Risk-minimizing hedging strategies for insurance payment processes. *Finance and Stochastics* **5**(4), pp. 419–446. DOI: 10.1007/s007800100041.

Møller, T. (2002). On Valuation and Risk Management at the Interface of Insurance and Finance. *British Actuarial Journal* **8**(4), pp. 787–827. DOI: 10.1017/S1357321700003913.

Møller, T. and M. Steffensen (2007). *Market-Valuation Methods in Life and Pension Insurance*. International Series on Actuarial Science. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9780511543289.

Norberg, R. (1989a). A Class of Conjugate Hierarchical Priors for Gammoid Likelihoods. *Scandinavian Actuarial Journal* **1989**(4), pp. 177–193. DOI: 10.1080/03461238.1989.10557353.

Norberg, R. (1989b). Experience Rating in Group Life Insurance. *Scandinavian Actuarial Journal* **1989**(4), pp. 194–224. DOI: 10.1080/03461238.1989.10557354.

Norberg, R. (1990). Payment Measures, Interest, and Discounting. *Scandinavian Actuarial Journal* **1990**, pp. 14–33. DOI: 10.1080/03461238.1990.10413870.

Norberg, R. (1991). Reserves in Life and Pension Insurance. *Scandinavian Actuarial Journal* **1991**, pp. 3–24. DOI: 10.1080/03461238.1991.10557357.

Norberg, R. (1992). Hattendorff's theorem and Thiele's differential equation generalized. *Scandinavian Actuarial Journal* **1992**, pp. 2–14. DOI: 10.1080/03461238.1992.10413894.

Norberg, R. (1995). A time-continuous Markov chain interest model with applications to insurance. *Applied Stochastic Models and Data Analysis* **11**(3), pp. 245–256. DOI: 10.1002/asm.3150110306.

Norberg, R. (1996). Addendum to Hattendorff's Theorem and Thiele's Differential Equation Generalized, SAJ 1992, 2–14. *Scandinavian Actuarial Journal* **1996**, pp. 50–53. DOI: 10.1080/03461238.1996.10413962.

Norberg, R. (1999). A theory of bonus in life insurance. *Finance and Stochastics* **3**, pp. 373–390. DOI: 10.1007/s007800050067.

Norberg, R. (2001). On Bonus and Bonus Prognoses in Life Insurance. *Scandinavian Actuarial Journal* **2001**(2), pp. 126–147. DOI: 10.1080/03461230152592773.

Norberg, R. (2010). Forward mortality and other vital rates – Are they the way forward? *Insurance: Mathematics and Economics* **47**, pp. 105–112. DOI: 10.1016/j.insmatheco.2010.07.002.

Norberg, R. (2013). Optimal hedging of demographic risk in life insurance. *Finance and Stochastics* **17**, 197–222. DOI: 10.1007/s00780-012-0182-3.

Organisation for Economic Co-operation and Development (2015). *Stocktaking of the tax treatment of funded private pension plans in OECD and EU countries (OECD).* https://www.oecd.org/tax/tax-treatment-funded-private-pension-plans-oecd-eu-countries.htm.

Perkowski, N. and J. Ruf (2015). Supermartingales as Radon–Nikodym densities and related measure extensions. *Annals of Probability* **43**(6), pp. 3133–3176. DOI: 10.1214/14-AOP956.

Ramlau-Hansen, H. (1991). Distribution of Surplus in Life Insurance. *ASTIN Bulletin* **21**, 57–71. DOI: 10.2143/AST.21.1.2005401.

Schweizer, M. (1994). Risk-Minimizing Hedging Strategies under Restricted Information. *Mathematical Finance* **4**(4), pp. 327–342. DOI: 10.1111/j.1467-9965.1994.tb00062.x.

Schweizer, M. (2001). "A Guided Tour through Quadratic Hedging Approaches". In: *Option Pricing, Interest Rates and Risk Management.* Ed. by E. Jouini, J. Cvitanić, and M. Musiela. Cambridge University Press, pp. 538–574. DOI: 10.1017/CBO9780511569708.016.

Schweizer, M. (2008). Local risk-minimization for multidimensional assets and payment streams. *Banach Center Publications* **83**, pp. 213–229. DOI: 10.4064/bc83-0-13.

Serfozo, R.F. (1971). Functions of Semi-Markov Processes. *SIAM Journal on Applied Mathematics* **20**(3), pp. 530–535. DOI: 10.1137/0120055.

Sokol, A. (2015). *Revisiting the forward equations for inhomogeneous semi-Markov processes.* arXiv: 1504.02955.

Steffensen, M. (2000). A no arbitrage approach to Thiele's differential equation. *Insurance: Mathematics and Economics* **27**(2), pp. 201–214. doi: 10.1016/S0167-6687(00)00048-2.

Steffensen, M. (2006). Surplus-linked life insurance. *Scandinavian Actuarial Journal* **2006**, pp. 1–22. doi: 10.1080/03461230500199533.

Van Der Gaag, N.L., G. Bijwaard, J. de Beer, and L. Bonneux (2015). A multistate model to project elderly disability in case of limited data. *Demographic Research* **32**(3), pp. 75–106. doi: 10.4054/DemRes.2015.32.3.

Vaupel, J.W., K.G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, pp. 439–454. doi: 10.2307/2061224.

Wong, W. H. (1986). Theory of Partial Likelihood. *Annals of Statistics* **14**, pp. 88–123. doi: 10.1214/aos/1176349844.