Matrix methods in multi-state life insurance

Jamaal Ahmad

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

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

PhD Thesis by:

Jamaal Ahmad
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
jamaal_ahmad@hotmail.com

Assessment committee:

Professor Thomas Mikosch (chairperson)
University of Copenhagen

Professor Hansjörg Albrecher
University of Lausanne, Switzerland

Professor Emeritus Søren Asmussen
Aarhus University, Denmark

Supervisor:

Professor Mogens Bladt
University of Copenhagen

Co-supervisor:

Professor Mogens Steffensen
University of Copenhagen

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen on 30 December 2022.

ISBN: 978-87-7125-066-4.

Abstract

This thesis considers matrix methods in multi-state life insurance, with an emphasis on techniques related to inhomogeneous phase-type distributions (IPH) and product integrals. We start out with developing an expectation-maximization (EM) algorithm for statistical estimation of general IPHs. Then we introduce a new class of multi-state models, the so-called aggregate Markov model, which allows for non-Markovian modeling with most of the analytical tractability of Markov chains preserved. Using techniques related to IPHs, we derive distributional properties, computational schemes for life insurance valuations with duration-dependent payments, and statistical estimation procedures based on the EM algorithm for general IPHs. Special attention is given to a case with a reset property, where the aggregate Markov model is semi-Markovian. We then move on and consider Markov chain interest rate models and show that bond prices are survival functions of IPHs. This allows for calibration via EM algorithms for phase-type distributions. Then we consider a multivariate payment process and derive higher order moments of its present value. Finally, we consider computation of market values of bonus payments in multi-state with-profit life insurance, where numerical procedures based on simulation of financial scenarios and classic analytical methods for insurance risk are developed.

Preface

This thesis has been prepared in 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 under the supervision of Professor Mogens Bladt (University of Copenhagen), while Professor Mogens Steffensen (University of Copenhagen) has acted as co-supervisor.

The main part of the thesis consists of an introduction followed by six chapters that are based on manuscripts written throughout the study period. Although the topics between the manuscripts are related, they appear as independent scientific contributions and should therefore also be read with this in mind. In particular, the notation slightly varies between chapters, and overviews of existing literature and concepts are sometimes repeated. The introduction serves the purpose of providing the relations between the chapters and their main contributions, thereby giving the reader an overall story behind the studies undertaken in the thesis.

This is it. After almost all of my adulthood so far has been spent at the H.C. Ørsted Institute on or around the Department of Mathematical Sciences at the University of Copenhagen, it now comes to an end with this thesis. I have grown as a person through these years, and the last years spent on matrix methods in life insurance could not have been a better way to end this immense journey of learning. It is my hope that the reader will enjoy this culmination at least as much as I have.

Acknowledgments

There is one person that stands out when it comes to the success of the project, and that is my principal supervisor Mogens Bladt. Your interest and commitment towards me and my project has been tremendous, and I have been very happy to have had the opportunity to learn from your extensive knowledge and experience. Thanks for being there, for your nature as a supervisor, and for all of our fruitful discussions by your blackboard. I could not have asked for a better time.

A special thanks also goes to my co-supervisor Mogens Steffensen. More than four years ago, when I came to you with my initial thoughts of going for the PhD, you were able to inspire me in a way that converted these thoughts into a serious path to take. It is fair to say that you played a huge part in making this possible from the beginning. Thank you for always making me think bigger.

Many thanks to the Department of Mathematical Sciences, University of Copenhagen, for awarding me one of your general PhD stipends. It has meant a lot to me, and I will be forever grateful for this recognition. Thanks to my colleagues at the section for Insurance and Economics for making such a nice research environment. I have enjoyed being surrounded by all the great and unique personalities that walks around the floors. Especially, I should like to thank Christian Furrer for our many talks, for always being able to change my mood - in both directions, and for providing this beautiful thesis template. Thanks to my office mates throughout the years: Alexander, Anna, Debbie, Gabriele, Jacob, Julie, Nena, and Oliver, for all the walks and talks, and for contributing to a nice everyday life on campus.

From March to June 2022, I had the opportunity to carry out my change of research environment by visiting the Department of Actuarial Science at University of Lausanne, Switzerland. I should like to thank Professor Hansjörg Albrecher for hosting me, and the rest of the department for a memorable time in Lausanne. I am very thankful to have been an integrated part of your professional and social activities. Especially, I thank my office mates Martin Bladt and Oscar Peralta for many fruitful discussions on common research interests. Thanks to the Department of Mathematical Sciences, University of Copenhagen, and 'Oberstløjtnant Max Nørgaard og Hustru Magda Nørgaards Fond' for jointly supporting this trip.

Thanks to my co-authors Martin Bladt, Mogens Bladt, Kristian Buchardt, and Christian Furrer for our collaborations over the years. I have learned a lot from you during the many hours spent together on the projects. It is safe to say that with great personalities comes great discussions.

Last but not least, from the bottom of my heart, I thank Cecilie for all the support throughout the years. Your faith in me has been out of this world. Thanks for always being there for me, listening to all my frustrations and joy, and for encouraging me to go for the PhD in the first place. None of this could have happened without you.

Jamaal Ahmad
Copenhagen, December 2022

List of papers

The main part of the thesis is based on the following manuscripts, which are submitted for publication or already published in international peer-reviewed journals:

Chapter 2:
Ahmad, J., M. Bladt, and M. Bladt (2022). Estimating absorption time distributions of general Markov jump processes. *arXiv:2207.11303*.

Chapter 3:
Ahmad, J., M. Bladt, and C. Furrer (2022). Aggregate Markov models in life insurance: properties and valuation. *arXiv:2212.03705*.

Chapter 4:
Ahmad, J. and M. Bladt (2022a). Aggregate Markov models in life insurance: estimation via the EM algorithm. *arXiv:2212.10661*.

Chapter 5:
Ahmad, J. and M. Bladt (2022b). Phase-type representations of stochastic interest rates with applications to life insurance. *arXiv:2207.11292*.

Chapter 6:
Ahmad, J. (2022). Multivariate higher order moments in multi-state life insurance. *Scandinavian Actuarial Journal* **2022**(5), pp. 399–420.

Chapter 7:
Ahmad, J., K. Buchardt, and C. Furrer (2022). Computation of bonus in multi-state life insurance. *ASTIN Bulletin* **52**(1), 291–331.

Summary

This thesis is mainly about matrix methods in multi-state life insurance. We start out with giving an introductory background and overview of the main contributions in Chapter 1, which then is followed by Chapters 2–7 where we present the manuscripts forming the main part of the thesis. A principal focus of the thesis, which constitutes the Chapters 2–5, is the application of inhomogeneous phase-type distributions in multi-state life insurance.

We begin the journey in Chapter 2 with the manuscript *Ahmad, Bladt, and Bladt (2022)* where we develop EM algorithms for statistical fitting of general IPHs. The EM algorithm is obtained using techniques known from statistical inference of time-inhomogeneous Markov jump processes based on their multivariate counting processes. We implement the algorithm and apply it to data in the case of piecewise constant transition rates, allowing for simplified estimation procedures based on Poisson regressions on a set of occurrences and exposures.

In Chapters 3–4, we then introduce a class of multi-state life insurance models we refer to as *aggregate Markov models*. They are constructed by adding unobservable sub-states, so-called *microstates*, to each biometric or behavioral state, thereby referring to the latter as *macrostates*. The idea is to obtain conditional sojourn time distributions that are IPH of general dimension; the classic Markov chain models have one-dimensional IPHs as conditional sojourn time distributions. In Chapter 3, which is based on the manuscript *Ahmad, Bladt, and Furrer (2022)*, we derive distributional properties as well as computational schemes for life insurance valuations of duration-dependent payments. The results reveal that aggregate Markov models can be highly non-Markovian. Throughout, we give special attention to a case, which we refer to as the *reset property*, where the aggregate Markov process is a time-inhomogeneous semi-Markov process.

Then, in Chapter 4, which is based on the manuscript *Ahmad and Bladt (2022a)*, we consider statistical estimation in the aggregate Markov model based on data of trajectories of the macrostate process. Using similar techniques as in Chapter 2, combined with the many distributional results derived in Chapter 3, we develop EM algorithms for the estimation of transition rates on the micro level. Also here,

we give special attention to the reset property case, especially since it allows us to use the EM algorithms of Chapter 2 as input. We implement the latter in the case of piecewise constant transition rates and apply it to data simulated from a time-inhomogeneous semi-Markovian disability model. The resulting estimated transition rates are also used in a numerical example in Chapter 3 to carry out life insurance valuations of a disability coverage with a waiting period.

In Chapter 5, which is based on the manuscript *Ahmad and Bladt (2022b)*, we consider stochastic interest rates following a time-inhomogeneous Markov jump process, the so-called Markov chain market. By deriving suitable product integral representations of the bond prices, we show that these are survival functions of IPHs. This allows us to fit (calibrate) the transition rates of the underlying Markov chain from observed bond prices using EM algorithms for phase-type distributions. We provide some numerical examples to illustrate this. Furthermore, we show how the model naturally integrates into existing matrix frameworks in multi-state life insurance, by providing product integral representations of reserves and higher order moments with stochastic interest rates on this form.

In Chapter 6, which is based on the paper *Ahmad (2022)*, we consider a multivariate payment process with components defined in terms of the same underlying time-inhomogeneous Markov jump process. We derive differential equations and product integral representations of higher order moments of the multivariate present value. This allows us to analyze joint effects between different product types in a general multi-state Markovian framework. Special attention is given to pairwise covariances and correlations between two product types, where results related to Hattendorff type of results for the variance are derived.

The thesis then ends with Chapter 7, which is based on the paper *Ahmad, Buchardt, and Furrer (2022)*. Here, the problem of computing the market value of bonus payments in multi-state with-profit life insurance is attended. We consider the bonus scheme known as additional benefits, where dividends are used to buy extra benefits to the insured. By assuming that dividends are affine in the number of additional benefits held, we derive differential equations that allows for computation of market values of bonus payments using a combination of simulating financial scenarios and classic analytical methods for insurance risk. We give special attention to the case where the number of additional benefits only depends on financial risk, which allows for simplified numerical procedures.

Resumé

Denne afhandling handler hovedsageligt om matrix metoder i flertilstandslivsforsikring. Vi starter ud med at give relevant baggrundsstof og en oversigt over hovedbidragene i Kapitel 1, som efterfølges af Kapitel 2–7, hvor vi præsenterer de manuskripter, der udgør kernen af afhandlingen. Et hovedfokus i afhandlingen, som udgør Kapitel 2–5, er anvendelsen af inhomogene fasetypefordelinger (IPH) i flertilstandslivsforsikring.

Vi begynder rejsen i Kapitel 2 med manuskriptet *Ahmad, Bladt og Bladt (2022)*, hvor vi udvikler EM algoritmer til statistisk estimation af generelle IPH'er. EM algoritmerne opnås ved hjælp af teknikker kendt fra statistisk inferens for tidsinhomogene Markov springprocesser baseret på deres flerdimensionelle tælleprocesser. Vi implementerer algoritmen og anvender den på data i tilfældet med stykkevis konstante intensiteter, hvilket giver mulighed for simplere estimations metoder baseret på Poisson regressioner på et sæt af antal hændelser og eksponeringer.

I Kapitel 3–4 introducerer vi derefter en klasse af flertilstandsmodeller vi omtaler som *aggregerede Markov modeller*. De er konstrueret ved at tilføje underliggende uobserverbare tilstande, såkaldte *mikrotilstande*, til hver biometrisk eller adfærdsmæssig tilstand, hvorved disse derfor omtales som *makrotilstande*. Ideen er at opnå betingede fordelinger for opholdstider der er IPH af generel dimension; de klassiske Markov modeller har en-dimensionelle IPH'er som betingede fordelinger for opholdstider. I Kapitel 3, som er baseret på manuskriptet *Ahmad, Bladt og Furrer (2022)*, udleder vi fordelingsmæssige egenskaber samt beregningsmetoder for værdiansættelse af livsforsikringsforpligtelser med varighedsafhængige betalinger. Resultaterne afslører, at aggregerede Markov modeller kan være særdeles ikke-Markovianske. Vi lægger særligt vægt på et specialtilfælde, som vi omtaler som *nulstillings egenskaben*, hvor den aggregerede Markov proces er en tidsinhomogen semi-Markov proces.

Derefter, i Kapitel 4, som er baseret på manuskriptet *Ahmad og Bladt (2022a)*, behandler vi statistisk estimation i den aggregerede Markov model baseret på data af stier for makro tilstandsprocessen. Ved at bruge lignende teknikker som i Kapitel 2, kombineret med de mange fordelingsresultater udledt i Kapitel 3, udvikler vi EM algoritmer til estimation af intensiteter på mikroniveau. Også her lægger vi særligt vægt på tilfældet med nulstillings egenskaben, især da det giver os mulighed for at bruge EM algoritmerne fra Kapitel 2 som input. Vi implementerer sidstnævnte i tilfældet med stykkevis konstante intensiteter og anvender den på data simuleret fra en tidsinhomogen semi-Markoviansk invalidemodel. De estimerede intensiteter vi opnår her bruges også i et numerisk eksempel i Kapitel 3 til at udføre værdiansættelse af en invalidedækning med karenstid.

I Kapitel 5, som er baseret på manuskriptet *Ahmad og Bladt (2022b)*, betragter vi stokastiske renter modelleret som tidsinhomogene Markov springprocesser, det såkaldte Markov-kæde marked. Ved at udlede passende produkt integral repræsentationer for obligationspriser, viser vi, at disse er overlevelsesfunktioner for IPH'er. Dette giver os mulighed for at estimere (kalibrere) intensiteterne for den underliggende Markov kæde fra observerede obligationspriser ved hjælp af EM algoritmer for fasetypefordelinger. Vi illustrerer dette gennem en række numeriske eksempler. Ydermere viser vi hvordan modellen naturligt integreres i eksisterende matrix-baserede modeller i flertilstandslivsforsikring, hvor vi udleder produkt integral repræsentationer for reserver og højere ordens momenter med stokastiske renter på denne form.

I Kapitel 6, som er baseret på artiklen *Ahmad (2022)*, betragter vi en flerdimensionel betalingsproces med komponenter defineret i termer af den samme underliggende tidsinhomogene Markov springproces. Vi udleder differentialligninger og produkt integral repræsentationer for højere ordens momenter af den flerdimensionelle nutidsværdi. Dette giver os mulighed for at analysere afhængigheder mellem forskellige produkttyper i en generel flertilstands Markoviansk ramme. Der lægges særlig vægt på kovariansen og korrelationen mellem to produkttyper, hvor resultater relateret til Hattendorff's resultater for variansen udledes.

Afhandlingen afsluttes derefter med Kapitel 7, som er baseret på artiklen *Ahmad, Buchardt og Furrer (2022)*. Her behandles problemet med at beregne markedsværdien af bonusbetalinger, bonuspotentialet, for gennemsnitsrente i flertilstandslivsforsikring. Vi betragter bonusordningen ydelsesopskrivning, hvor dividender bruges til at købe ekstra ydelser til den forsikrede. Ved at antage, at dividenderne er affine i antallet af ydelser tilkøbt, udleder vi differentialligninger der giver mulighed for beregning af markedsværdien af bonusbetalinger ved hjælp af en kombination af simulering af finansielle scenarier og klassiske analytiske metoder for forsikringsrisiko. Vi lægger særligt vægt på det tilfælde, hvor antallet af tilkøbte ydelser kun afhænger af finansiel risiko, hvilket giver mulighed for forenklede numeriske procedurer.

Contents

Chapter 1

Introduction

This thesis is mainly about matrix methods in multi-state life insurance, with an emphasis on techniques related to inhomogeneous phase-type distributions and product integrals. This introduction sets the scene for the studies undertaken. We start out with giving some background on existing models and methods in the literature on multi-state life insurance that are relevant for the present thesis. Based on this, we end the chapter by motivating the studies carried out in the subsequent chapters along with an overview of the main contributions.

1.1 Background

We now provide some background on multi-state modeling in life insurance. In Subsection 1.1.1, we present the classic Markov chain models, where special attention is paid to recent developments on matrix representations within this framework. Subsection 1.1.2 then considers semi-Markov models including a discussion on its added complexities relative to the Markov models. Finally, in Subsection 1.1.3, we consider the notion of inhomogeneous phase-type distributions and discuss modeling potentials in relation to the previous subsections when having these as building blocks.

1.1.1 Markov chain models

Markov chain models dates back to at least Hoem (1969a) and Norberg (1991), and are the most classic and popular approaches to multi-state life insurance modeling. It provides a unifying framework to model insurance contracts related to different kinds of life and health events in a tractable and computationally simple way. The model therefore naturally plays a predominant role throughout the thesis as a baseline model we extend in various directions. We give an overview of the framework in this subsection.

The states of the insured is governed by a time-inhomogeneous Markov jump process $Z = \{Z(t)\}_{t\geq 0}$ taking values on a finite state space $\mathcal{J} = \{1, \ldots, J\}$, $J \in \mathbb{N}$, indicating biometric or behavioral states of the insured. It is assumed to admit suitably regular transition rates $t \mapsto \mu_{ij}(t)$, $i, j \in \mathcal{J}$, $j \neq i$, with $\mu_{i\bullet}(t) = \sum_{\substack{j\in\mathcal{J} \\ j\neq i}} \mu_{ij}(t)$, implying that the transition probabilities

$$p_{ij}(t,s) = \mathbb{P}(\, Z(s) = j \,|\, Z(t) = i)$$

satisfy Kolmogorov's forward and backward differential equations:

$$\begin{aligned} \frac{\partial}{\partial s} p_{ij}(t,s) &= \sum_{k:k\neq j} p_{ik}(t,s)\mu_{kj}(s) - p_{ij}(t,s)\mu_{j\bullet}(s), \qquad p_{ij}(t,t) = 1_{(i=j)}, \\ \frac{\partial}{\partial t} p_{ij}(t,s) &= -\sum_{k:k\neq i} \mu_{ik}(t)p_{kj}(t,s) + \mu_{i\bullet}(t)p_{ij}(t,s), \quad\; p_{ij}(s,s) = 1_{(i=j)}. \end{aligned} \tag{1.1.1}$$

The life insurance contract is then modeled by a payment process $B = \{B(t)\}_{t\geq 0}$ giving accumulated benefits less premiums. It is assumed to consist of payment rates during sojourn states and payments upon transition between states, thus taking the form

$$\begin{aligned} \mathrm{d}B(t) &= \sum_{j\in\mathcal{J}} \left(1_{(Z(t)=j)} b_j(t)\,\mathrm{d}t + \sum_{\substack{k\in\mathcal{J} \\ k\neq j}} b_{jk}(t)\,\mathrm{d}N_{jk}(t) \right), \\ B(0) &\in \mathbb{R}, \end{aligned} \tag{1.1.2}$$

where $b_j(t)$ and $b_{jk}(t)$ are suitably regular deterministic payment functions, and N is the multivariate counting process associated to Z, with components $N_{jk} = \{N_{jk}(t)\}_{t\geq 0}$, $j, k \in \mathcal{J}$, $k \neq j$, given by

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

For simplicity, we may assume a maximal contract time $T > 0$ such that $b_j(t) = b_{jk}(t) = 0$ for all $t > T$; this could e.g. be a maximal living age of the insured.

For valuation of the life insurance liabilities, the prospective reserve is the key. It is given as the expected present value of future payments given the available information. Assuming a deterministic and suitably regular interest rate $r(t)$, it then reads

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

where $\mathcal{F} = \{\mathcal{F}(t)\}_{t\geq 0}$ denotes the natural filtration generated by Z, representing the available information.

The Markov assumption on Z together with the assumption that the payments at some time only depends on Z through its value at that time, that is, that the payment functions b_j and b_{jk} are deterministic, gives the state-wise counterparts $V(t) = V_{Z(t)}(t)$, where

$$V_i(t) = \mathbb{E}\left[\int_t^T e^{-\int_t^s r(v)\,\mathrm{d}v}\,\mathrm{d}B(s) \,\middle|\, Z(t) = i \right], \quad i \in \mathcal{J},$$

are state-wise prospective reserves. They satisfy the celebrated Thiele's differential equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} V_i(t) &= r(t)V_i(t) - b_i(t) - \sum_{\substack{j\in\mathcal{J}\\ j\neq i}} \mu_{ij}(t)(b_{ij}(t) + V_j(t) - V_i(t)), \\ V_i(T) &= 0. \end{aligned} \tag{1.1.3}$$

Computation of the state-wise prospective reserves can then be carried out by solving the backward system of differential equations (1.1.3), which then provides the reserves at all time points between an initial time and the maximal contract time.

Several extensions of the classic multi-state version of Thiele's differential equation (1.1.3) has been considered in the literature. Within Markov chain models, this, e.g., includes higher order moments of present values in Norberg (1995b) and inclusion of stochastic interest following a Markov jump process in Norberg (1995a, 2003), but also Hattendorff type of differential equations for the variance of present values have been considered in Ramlau-Hansen (1988).

In recent years, there has been an increased interest towards representing the prospective reserve in terms of so-called expected accumulated cash flows. Following Buchardt, Furrer, and Steffensen (2019, Definition 2.2), they are given by, for a valuation time $t \geq 0$,

$$A(t,s) = \mathbb{E}[\,B(s) - B(t)\,|\,\mathcal{F}(t)], \quad s \geq t, \tag{1.1.4}$$

which then gives the following integral expression for the prospective reserve:

$$V(t) = \int_t^T e^{-\int_t^s r(v)\,\mathrm{d}v} A(t,\,\mathrm{d}s).$$

In the Markov chain model, it was already carried out by Buchardt and Møller (2015). Here, the expected accumulated cash flow takes the form $A(t,s) = A_{Z(t)}(t,s)$, with the state-wise counterparts satisfying

$$\begin{aligned} V_i(t) &= \int_t^T e^{-\int_t^s r(v)\,\mathrm{d}v} A_i(t,\,\mathrm{d}s), \\ A_i(t,\,\mathrm{d}s) &= \sum_{j\in\mathcal{J}} p_{ij}(t,s)\Big(b_j(s) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(s)\mu_{jk}(s)\Big)\mathrm{d}s. \end{aligned} \tag{1.1.5}$$

Computation of the prospective reserve via the expected accumulated cash flow thus requires a computation of the transition probabilities $p_{ij}(t,\cdot)$ via Kolmogorov's forward differential equations. This gives the reserve at a single time t, as opposed to Thiele's differential equation (1.1.3) which gives the reserve at all time points. However, if one is interested in computing the reserve for different interest rate levels to asses sensitivities, e.g. in the context of hedging interest rate risk when it is stochastic, the cash flow method (1.1.5) provides an efficient alternative.

Matrix representations and product integrals

In the context of survival and event history analysis, and thus also Markov processes, there has been occasional use of so-called product integrals as a tool to describe solutions to differential equations of Kolmogorov type, see, e.g., Johansen (1986) and Gill and Johansen (1990) for a survey. It was recently put into the multi-state Markovian life insurance context by Bladt, Asmussen, and Steffensen (2020), where they derive product integral representations of reserves using the close relations between Thiele's and Kolmogorov's differential equations. This allows for a more compact and direct treatment of Markov chain models, especially in relation to numerical implementation.

We outline the concept in the following. Consider the transition intensities $\mu_{ij}(t)$ of the Markov jump process Z on matrix form as $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j\in\mathcal{J}}$, where $\mu_{ii}(t) = -\mu_{i\bullet}(t)$, and likewise for the corresponding transition probabilities $\boldsymbol{P}(t,s) = \{p_{ij}(t,s)\}_{i,j\in\mathcal{J}}$. Then Kolmogorov's forward and backward differential equations can be written compactly on the form

$$\begin{aligned} \frac{\partial}{\partial s}\boldsymbol{P}(t,s) &= \boldsymbol{P}(t,s)\boldsymbol{M}(s), \qquad \boldsymbol{P}(t,t) = \boldsymbol{I}, \\ \frac{\partial}{\partial t}\boldsymbol{P}(t,s) &= -\boldsymbol{M}(t)\boldsymbol{P}(t,s), \quad \boldsymbol{P}(s,s) = \boldsymbol{I}. \end{aligned} \tag{1.1.6}$$

The solution to such a system of differential equations is referred to as the product integral of $\boldsymbol{M}$ from t to s, and we write

$$\boldsymbol{P}(t,s) = \prod_t^s (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x). \tag{1.1.7}$$

Here, the concept is not restricted to intensity matrices but instead holds for all matrix functions satisfying (1.1.6) (whenever a solution exist). Treating the solution as an object of its own allows us to draw upon its many properties for further derivations and manipulations.

It is then shown in Bladt, Asmussen, and Steffensen (2020) that the Thiele type of differential equations closely resembles (1.1.6)–(1.1.7). The key is the introduction of so-called partial state-wise prospective reserves

$$V_{ij}(t) = \mathbb{E}\left[1_{(Z(T)=j)} \int_t^T e^{-\int_t^s r(v)\,\mathrm{d}v}\, \mathrm{d}B(s) \,\middle|\, Z(t) = i \right], \qquad (1.1.8)$$
$$\boldsymbol{V}(t) = \{V_{ij}(t)\}_{i,j\in\mathcal{J}},$$

which generalizes the transition probabilities in the sense that if we take the 0'th moment of the present value, we arrive at the transition probabilities; the reserves are first order moments. Then, by introducing the reward matrix function $\boldsymbol{R}(t)$, given by

$$\boldsymbol{R}(t) = \boldsymbol{M}(t) \bullet \boldsymbol{B}(t) + \boldsymbol{\Delta}(\boldsymbol{b}(t)),$$
$$\boldsymbol{B}(t) = \{b_{ij}(t)\}_{i,j\in\mathcal{J}},$$
$$\boldsymbol{b}(t) = (b_1(t), \ldots, b_J(t))',$$

where $\bullet$ denotes the Schur product, that is $(\boldsymbol{A} \bullet \boldsymbol{B})_{ij} = A_{ij}B_{ij}$, and $\boldsymbol{\Delta}(\boldsymbol{b})$ is a diagonal matrix with the vector $\boldsymbol{b}$ as diagonal, the following product integral representation is achieved

$$\prod_t^T \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{R}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix} \mathrm{d}x \right) = \begin{pmatrix} e^{-\int_t^T r(v)\,\mathrm{d}v} \boldsymbol{P}(t,T) & \boldsymbol{V}(t) \\ \boldsymbol{0} & \boldsymbol{P}(t,T) \end{pmatrix}.$$

This follows from extending results in Van Loan (1978) from matrix exponentials to product integrals, see also Bladt, Asmussen, and Steffensen (2020, Lemma 2).

The block structure of the matrix put up on the left hand side gives a compact and unifying treatment of transition probabilities and reserves in a single notion. Further results for higher order moments are derived in similar fashion; one simply adds a suitable block row for each moment. We refer to Bladt, Asmussen, and Steffensen (2020, Theorem 5) for the details.

Although matrix representations of the expected accumulated cash flow (1.1.4) is not directly mentioned in Bladt, Asmussen, and Steffensen (2020), they are readily obtained via the reward matrix function $\boldsymbol{R}(t)$ as follows. Defining the vector containing the state-wise counterparts,

$$\boldsymbol{A}(t,s) = (A_1(t,s), \ldots, A_J(t,s))',$$

we get by (1.1.5),

$$\boldsymbol{A}(t,\,\mathrm{d}s) = \boldsymbol{P}(t,s)\boldsymbol{R}(s)\boldsymbol{1}_J\,\mathrm{d}s.$$

In the following subsections, we consider a number of alternatives to the classic Markov chain models that allows for added flexibility, and discuss their relations to the properties of Markov models considered in this subsection.

1.1.2 Semi-Markov models

The Markov chain model enjoys a number of technical and computational advantages, which play a fundamental role in its popularity. However, it suffers from not being able to capture duration effects, which evidently appear in the context of multi-state life insurance. This has motivated the need for more sophisticated models, and the time-inhomogeneous semi-Markov model has seen considerable attention over the years, see, e.g., Hoem (1972), Helwich (2008), Christiansen (2012), and Buchardt, Møller, and Schmidt (2015).

In the semi-Markov model one instead assumes that (Z, U) is Markovian, where U is the process giving the duration since the last jump in Z:

$$U(t) = \sup\big\{s \in [0,t] : Z(u) = Z(t) \text{ for all } u \in [t-s,t]\big\}. \tag{1.1.9}$$

The model is then described by transition rates $\mu_{ij}(t,u)$ depending on both the absolute time t and the duration u since the last jump. In addition to this, the payments also depend on the duration process U, such that it takes the form

$$\begin{aligned} \mathrm{d}B(t) &= \sum_{j\in\mathcal{J}} \bigg(1_{(Z(t)=j)} b_j(t, U(t))\,\mathrm{d}t + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(t, U(t-))\,\mathrm{d}N_{jk}(t) \bigg), \\ B(0) &\in \mathbb{R}, \end{aligned} \tag{1.1.10}$$

where $b_j(t,u)$ and $b_{jk}(t,u)$ are suitably regular deterministic payment functions depending on both time and duration.

The expected accumulated cash flow now takes the form $A(t,s) = A_{Z(t),U(t)}(t,s)$, and similarly for the prospective reserve, $V(t) = V_{Z(t),U(t)}(t)$, with

$$\begin{aligned} V_{i,u}(t) &= \int_t^T e^{-\int_t^s r(v)\,\mathrm{d}v} A_{i,u}(t,\,\mathrm{d}s), \\ A_{i,u}(t,\,\mathrm{d}s) &= \sum_{j\in\mathcal{J}} \int_0^{u+s-t} p_{ij}(t,u,s,\,\mathrm{d}z) \bigg(b_j(s,z) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(s,z)\mu_{jk}(s,z) \bigg) \mathrm{d}s, \end{aligned}$$

where the transition probabilities p are given by

$$p_{ij}(t,u,s,z) = \mathbb{P}(\, Z(s) = j, U(s) \leq z \,|\, Z(t) = i, U(t) = u).$$

The transition probabilities can be calculated by solving the forward integro-differential equations of Buchardt, Møller, and Schmidt (2015, Theorem 3.1), which corresponds to Kolmogorov's forward differential equations for the semi-Markovian case. Similarly, there exist partial differential equations for the prospective reserve itself, cf., e.g., Helwich (2008, Theorem 4.11) or Christiansen (2012, (3.9)), which suitably extends Thiele's differential equation to the semi-Markovian case.

The inclusion of duration effects from the semi-Markov model adds a significant layer to the computational complexity compared to the Markovian case of Subsection

1.1.1. Indeed, it now requires computations on a two-dimensional grid of time and duration, while a one-dimensional grid of time suffices for Markov chain models. In particular, the aforementioned matrix representations and product integral representations of Bladt, Asmussen, and Steffensen (2020) are, unfortunately, not applicable for semi-Markov models, since these techniques relies on ordinary differential equations.

In the following subsection, we consider a class of distributions that plays an important role throughout the thesis, which in particular aims at carrying over tools from Markovian modeling to semi-Markov models.

1.1.3 Inhomogeneous phase-type distributions

Phase-type distributions (PH), which are defined as absorption times of time-homogeneous Markov jump processes, have a long history of extensive use in applied probability. They are dense in the class of distributions on the positive reals, in the sense of weak convergence as the number of phases tends to infinity, and often lead to explicit solutions to complex problems due to their inherit tractability from Markov chains. For a comprehensive survey on PH distributions, see Bladt and Nielsen (2017).

In this subsection, we focus on the more recent developments made by Albrecher and Bladt (2019), where they introduce the class of inhomogeneous phase-type distributions (IPH). It is defined as follows. Consider a time-inhomogeneous Markov jump process $X = \{X(t)\}_{t\geq 0}$ taking values on the finite state space $\mathcal{J} = \{1, \ldots, J-1, J\}$, where the states $\{1, \ldots, J-1\}$ are transient and J is absorbing. Denote with $(\boldsymbol{\pi}, 0)$ the initial distribution of X, and $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j\in\mathcal{J}}$ the transition intensity matrix function of X. It is then on the form

$$\boldsymbol{M}(t) = \begin{pmatrix} \boldsymbol{T}(t) & \boldsymbol{t}(t) \\ 0 & 0 \end{pmatrix}, \tag{1.1.11}$$

where $\boldsymbol{T}(t)$ is a sub-intensity matrix function consisting of transition rates between the transient states and $\boldsymbol{t}(t) = -\boldsymbol{T}(t)\mathbf{1}_J$ is a column vector of transition rates to the absorbing state, the so-called exit rate vector function. The time until absorption, given by

$$\tau = \inf\{t \geq 0 \,:\, X(t) = J\},$$

is then said to be an inhomogeneous phase-type distribution with representation $(\boldsymbol{\pi}, \boldsymbol{T})$, and we write $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T})$.

The density $f(x)$ and distribution function $F(x)$ of τ are then obtained via product integrals of the sub-intensity matrix function $\boldsymbol{T}(t)$:

$$\begin{aligned} f(x) &= \boldsymbol{\pi} \prod_0^x (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{t}(x), \\ F(x) &= 1 - \boldsymbol{\pi} \prod_0^x (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{1}_J. \end{aligned} \tag{1.1.12}$$

An important feature of IPH distributions is that the overshoot of an IPH distribution is again IPH-distributed. This follows from (1.1.12), since then

$$\mathbb{P}(\tau > s + t \mid \tau > s) = \frac{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)}{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J} \prod_s^t (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J, \tag{1.1.13}$$

which shows that

$$\tau - s \mid \tau > s \sim \mathrm{IPH}(\boldsymbol{\alpha}(s), \boldsymbol{T}(s + \bullet)),$$

where $\boldsymbol{\alpha}(s)$ is given by

$$\boldsymbol{\alpha}(s) = \frac{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)}{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J}.$$

This property plays an important role throughout the thesis, as it implies a non memory-less property of IPH distributions. In other words, it reveals that they may be used to model duration effects even though they are defined in terms of Markov processes.

1.2 Overview of the thesis and main contributions

We end the introduction by giving an overview of the remaining chapters, which constitute the main part of the thesis and thus also contains the main contributions. Each chapter is based on independent manuscripts and should therefore also be read with this in mind. In particular, notation slightly varies between chapters, and since most of the manuscripts are related, overviews of existing literature and concepts are sometimes repeated between the chapters. While the following subsections intend to highlight the main contents of each chapter, the focal point of the exposition is to provide the necessary relations between the chapters, thereby giving the reader an overall story of the thesis. These relations are related to the background given in Section 1.1.

1.2.1 Inhomogeneous phase-type distributions in multi-state life insurance

Phase-type distributions (PH) have not seen substantial use in the context of multi-state life insurance so far. This is perhaps due to the time-homogeneity of the underlying Markov chain, which is not suitable in setups with strong age dependencies. However, with the recent introduction of inhomogeneous phase-type distributions (IPH) in Albrecher and Bladt (2019), as absorption times of time-inhomogeneous Markov jump processes, it now makes it very natural to consider applications within multi-state life insurance. This is a principal focus of the thesis and the overall theme of Chapters 2-5.

Estimation of general IPHs

To start of using IPHs in multi-state life insurance, we first develop tools for fitting them to data. This is the content of Chapter 2, where we present the manuscript *Ahmad, Bladt, and Bladt (2022).* This will later on allow us to estimate parameters in models where we use IPHs as building blocks. We tackle it as an incomplete data problem where only absorption times of time-inhomogeneous Markov jump processes are observed. An expectation-maximization (EM) algorithm is then employed to estimate parameters of the underlying Markov chain.

Fitting IPHs to data has earlier on been carried out by Albrecher, Bladt, and Yslas (2022), but only within a sub-class where the sub-intensity matrix function $\boldsymbol{T}(t)$ in (1.1.11) is of the form $\boldsymbol{T}(t) = \lambda(t)\boldsymbol{T}$, for a scalar function λ and sub-intensity matrix $\boldsymbol{T}$. This implies commuting sub-intensity matrix functions across time, and links to the conventional PH distributions (see Theorem 2.8 in Albrecher and Bladt, 2019) can then be used to develop EM algorithms where the classic EM algorithm of Asmussen, Nerman, and Olsson (1996) can be used as input.

In the multi-state life insurance context, however, we do not wish to restrict ourselves to this sub-class, as it limits the time-inhomogeneity needed to fully capture age dependencies. We therefore extend Albrecher, Bladt, and Yslas (2022) to general IPHs. Instead of taking PH distributions as starting point, which no longer is possible, we take as starting point Andersen et al. (1993), where general methods for statistical inference of time-inhomogeneous Markov jump processes based on their multivariate counting processes are available.

While we develop an EM algorithm for the general case, we propose a reduction to the case where the sub-intensity matrix function $\boldsymbol{T}(t)$ is piecewise constant on the form

$$\boldsymbol{T}(t) = \boldsymbol{T}_k = \left\{\mu_{ij}^k\right\}_{i,j\in\mathcal{J}}, \quad t \in (s_{k-1}, s_k], \quad k = 1, \dots, K,$$

for a suitable grid $s_0 = 0 < s_1 < \cdots < s_{K-1} < \infty = s_K$, and where we emphasize that the sub-intensity matrices $\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K$ in general do not commute.

Piecewise constant transition rates allows us to carry out simplified estimation procedures based on aggregated occurrences and exposures in the different time intervals, where Poisson regressions akin to those of Aalen, Borgan, and Gjessing (2008, Section 5) become available. Furthermore, general considerations via product integrals reduce to products of matrix exponentials, which computationally are more viable. We implement the EM algorithm in the piecewise constant case and show some numerical examples of mortality modeling of Danish lifetimes as well as fitting to theoretical distributions, illustrating the strength of our approach in these kinds of problems.

The example of mortality modeling, where we fit IPHs to lifetimes, can be seen as a prelude to their application in multi-state models, which is the theme of the following two chapters.

Aggregate Markov models

In Chapters 3-4, we introduce a class of multi-state models which we refer to as *aggregate Markov models*. The idea behind this class of models is as follows. Going back to the Markov chain models of Subsection 1.1.1, introduce the jump times $(T_n)_{n \in \mathbb{N}_0}$ of Z, where $T_0 = 0$. It then holds

$$\mathbb{P}(T_{n+1} > t \,|\, T_0, Z(T_0), T_1, Z(T_1), \ldots, T_n, Z(T_n) = j) = e^{\int_{T_n}^t \mu_{jj}(x)\,\mathrm{d}x}, \quad t \geq T_n,$$

which means that

$$T_{n+1} - T_n \,\Big|\, \big(T_i, Z(T_i)\big)_{i=0}^n \sim \mathrm{IPH}\big(1, \mu_{Z(T_n)Z(T_n)}(T_n + \bullet)\big).$$

In other words, sojourn time distributions in the Markov chain model follow one-dimensional IPH distributions independent of past sojourn times and transitions. Aggregate Markov models then extend to sojourn times admitting conditional IPH distributions of general dimension. This allows for added flexibility, such as duration dependence, while still retaining the analytical tractability from Markov chains.

The desired structure is obtained by adding so-called *microstates* to each biometric or behavioral state, which we then refer to as *macrostates*. This results in a two-dimensional state space

$$E = \{\boldsymbol{j} = (j, \widetilde{j}) : j \in \mathcal{J}, \widetilde{j} \in \{1, 2, ..., d_j\}\},$$

where $d_j \geq 1$, $j \in \mathcal{J}$, is the number of microstates assigned macrostate j.

The aggregate Markov model for the states of the insured Z is then given as follows. Introduce a time-inhomogeneous Markov jump process $\boldsymbol{X} = \{\boldsymbol{X}(t)\}_{t \geq 0} = \{(X_1(t), X_2(t))\}_{t \geq 0}$ taking values on E with transition intensity matrix function

$\boldsymbol{M}(t)$. Then $Z(t) = X_1(t)$ keeps track of the macrostate, while $X_2(t)$ identifies the current microstate contingent on the state of $X_1(t)$.

The transition intensity matrix function $\boldsymbol{M}(t)$ can be written on the following block form:

$$\boldsymbol{M}(t) = \begin{pmatrix} \boldsymbol{M}_{11}(t) & \boldsymbol{M}_{12}(t) & \cdots & \boldsymbol{M}_{1J}(t) \\ \boldsymbol{M}_{21}(t) & \boldsymbol{M}_{22}(t) & \cdots & \boldsymbol{M}_{2J}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{M}_{J1}(t) & \boldsymbol{M}_{J2}(t) & \cdots & \boldsymbol{M}_{JJ}(t) \end{pmatrix}, \tag{1.2.1}$$

where $\boldsymbol{M}_{jj}(t)$ are sub-intensity matrix functions of dimension $d_j \times d_j$ providing transition rates within macrostates, and $\boldsymbol{M}_{jk}(t)$ are non-negative matrix functions of dimension $d_j \times d_k$ providing transition rates between macrostates. Computations in aggregate Markov models are then, to a large extent, carried out using the block matrix functions $\boldsymbol{M}_{jk}(t)$ as building blocks, so to speak. This is in contrast to Markov chain models, where computations are based on transition rates, which can be viewed as 1×1 block matrices in (1.2.1).

In Chapter 3, which is based on the manuscript *Ahmad, Bladt, and Furrer (2022)*, we derive distributional properties of the macrostate process Z as well as computational schemes for life insurance valuation elements (prospective reserves and expected accumulated cash flows) with duration-dependent payments on the form (1.1.10). The distributional results reveal that aggregate Markov models can be highly non-Markovian. Throughout, we pay special attention to a case, which we refer to as the *reset property*, where the block matrix functions $\boldsymbol{M}_{jk}$, $j \neq k$, are rank one matrices of the following form

$$\boldsymbol{M}_{jk}(t) = \boldsymbol{\beta}_{jk}(t)\boldsymbol{\pi}_k(t), \tag{1.2.2}$$

where $\boldsymbol{\beta}_{jk}(t)$ is a d_j-dimensional non-negative column vector function and $\boldsymbol{\pi}_k(t)$ is a d_k-dimensional non-negative row vector function with $\boldsymbol{\pi}_k(t)\mathbf{1}_{d_k} = 1$. Here we show that Z admits a specific time-inhomogeneous semi-Markovian structure. Our main example is the disability model depicted in Figure 1.1 below, where we obtain duration effects regarding transitions out of the disabled state.

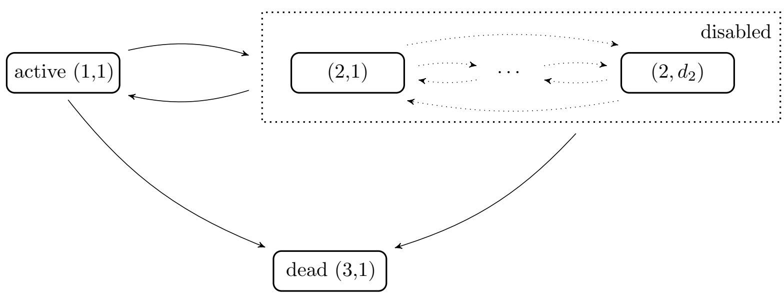

Figure 1.1: *Disability model with d_2 unobservable disability microstates.*

In Chapter 4, which is based on the manuscript *Ahmad and Bladt (2022a)*, we then consider statistical estimation in the aggregate Markov model based on independent observations of paths of the macrostate process. Like in Chapter 2, this is tackled as an incomplete data problem with respect to the underlying macro-micro state process $\boldsymbol{X}$, and EM algorithms are developed to estimate transition rates on the micro level. Here, the many distributional results derived in Chapter 3 are drawn upon. While we develop an EM algorithm for the general aggregate Markov model, we focus on the case where the reset property (1.2.2) is satisfied, since this case allows us to, more or less, use EM algorithms developed in Chapter 2 as input.

We implement the EM algorithm in the case where the reset property is satisfied along with piecewise constant transition rates, thereby being able to draw upon the implemented algorithm in Chapter 2. We provide a numerical example where we fit the model of Figure 1.1 to macro data simulated from a time-inhomogeneous semi-Markovian disability model employed by a large Danish life insurance company; the model is reported to and published by the Danish Financial Supervisory Authority. The resulting estimates are also used in a numerical example in Chapter 3 to calculate expected accumulated cash flows and prospective reserves for a disability coverage with a waiting period. Thus, the numerical examples of the two chapters intend to show how the aggregate Markov model with the reset property suitably approximates semi-Markov models in multi-state life insurance, both regarding the transition rates as well as corresponding valuation elements.

The Markov chain interest rate model

In Chapter 5, which is based on the manuscript *Ahmad and Bladt (2022b)*, we consider a somewhat different application of IPHs in multi-state life insurance. Going back to the classic Markov chain models of Subsection 1.1.1, we consider stochastic interest rates following the Markov chain itself, i.e.

$$r(u) = r_{Z(u)}(u), \tag{1.2.3}$$

for suitably regular deterministic functions $r_i(u)$, $i \in \mathcal{J}$. These type of models are introduced in Norberg (1995a, 2003) as the Markov chain market. Though well-studied there, a key result in our work is that if all $r_i(u) \geq 0$, then the bond price

$$B(t,T) = \mathbb{E}\Big[\mathrm{e}^{-\int_t^T r_{Z(u)}(u)\,\mathrm{d}u} \,\Big|\, \mathcal{F}(t)\Big], \quad 0 \leq t \leq T, \tag{1.2.4}$$

is the survival function of an IPH distribution (for fixed t). We also provide a similar (but scaled) relation in the presence of negative interest rates, with the only requirement that they are bounded from below.

While the functional form of the bond price (1.2.4) was noted already in Norberg (2003, (3.17)), its relation to phase–type theory was not mentioned, and its potential was not further explored. By doing so, it enables us to use the extensive toolbox provided by this class of distributions. In particular, we are able to fit (calibrate) the transition rates of $\{Z(u)\}_{u\geq 0}$ from the observed bond prices by using a maximum likelihood approach based on EM algorithms of Asmussen, Nerman, and Olsson (1996), Albrecher, Bladt, and Yslas (2022), and Ahmad, Bladt, and Bladt (2022). We provide a series of numerical examples to illustrate this.

In the context of multi-state life insurance, the main advantage of the model (1.2.3) is that it can be wholly incorporated into Thiele and Hattendorff type of differential equations for reserves and higher order moments; see Norberg (1995a,b, 2003). Following the partial state-wise reserves of (1.1.8), we introduce partial state-wise bond prices, given by

$$\begin{aligned} d_{ij}(t,T) &= \mathbb{E}\Big[1_{(Z(T)=j)}\mathrm{e}^{-\int_t^T r_{Z(u)}(u)\,\mathrm{d}u}\,\Big|\,Z(t)=i\Big], \\ \boldsymbol{D}(t,T) &= \{d_{ij}(t,T)\}_{i,j\in\mathcal{J}}. \end{aligned} \tag{1.2.5}$$

Providing suitable product integral representations of $\boldsymbol{D}(t,T)$, we then show how it naturally integrates into the matrix framework of Bladt, Asmussen, and Steffensen (2020). Essentially, one simply uses the vector $\boldsymbol{r}(t) = (r_1(t),\dots,r_J(t))'$ of the different interest rate levels, and computes

$$\prod_t^T\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{M}(x)-\boldsymbol{\Delta}(\boldsymbol{r}(x)) & \boldsymbol{R}(x)\\ \boldsymbol{0} & \boldsymbol{M}(x)\end{pmatrix}\mathrm{d}x\right)=\begin{pmatrix}\boldsymbol{D}(t,T) & \boldsymbol{V}(t)\\ \boldsymbol{0} & \boldsymbol{P}(t,T)\end{pmatrix}$$

to obtain relevant prospective reserves. This is closely related to the product integral representation in Subsection 1.1.1 for deterministic interest rates. Analogous results are obtained for higher order moments; we show a numerical example where we use these to approximate the density and distribution function of the present value using Gram-Charlier expansions.

1.2.2 Multivariate higher order moments

In Chapter 6, which is based on the paper *Ahmad (2022)*, we derive matrix representations of higher order moments in a setup where we consider a multivariate payment process. The components of the payment process are defined in terms of the same underlying Markov process Z. This allows us to analyze joint effects between different product types in a general multi-state Markovian framework, and it extends the differential equations of Norberg (1995b) as well as the matrix framework of Bladt, Asmussen, and Steffensen (2020) to multivariate payment processes.

The main example we think of is the computation of covariances and correlations between two product types, whereby the product moment between two present values is needed. Here, the product integral representation of Subsection 1.1.1

naturally extends, so that now it is the matrix

$$\boldsymbol{F}_{\boldsymbol{U}}^{(1,1)}(x) = \begin{pmatrix} \boldsymbol{M}(x) - 2r(x)\boldsymbol{I} & \boldsymbol{R}_2(x) & \boldsymbol{R}_1(x) & \boldsymbol{C}^{(1,1)}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{R}_1(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{R}_2(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix}$$

that needs to be product integrated. Here, $\boldsymbol{R}_1(x)$ and $\boldsymbol{R}_2(x)$ are reward matrix functions for the two product types in question, and $\boldsymbol{C}^{(1,1)}(x) = \boldsymbol{M}(x) \bullet \boldsymbol{B}_1(x) \bullet \boldsymbol{B}_2(x)$, where $\boldsymbol{B}_1(x)$ and $\boldsymbol{B}_2(x)$ contains transition payments for the two product types. The resulting product integral,

$$\prod_t^T \left(\boldsymbol{I} + \boldsymbol{F}_{\boldsymbol{U}}^{(1,1)}(x)\,\mathrm{d}x\right) = \begin{pmatrix} * & * & * & \boldsymbol{V}^{(1,1)}(t) \\ * & * & * & \boldsymbol{V}^{(1,0)}(t) \\ * & * & * & \boldsymbol{V}^{(0,1)}(t) \\ * & * & * & \boldsymbol{P}(t,T) \end{pmatrix},$$

then gives the required product moment $\boldsymbol{V}^{(1,1)}(t)$, as well as the reserve for the individual product types, $\boldsymbol{V}^{(1,0)}(t)$ and $\boldsymbol{V}^{(0,1)}(t)$. From this, further results for the covariance are derived, which naturally relates to Hattendorff type of results for the variance.

1.2.3 Computing the market value of bonus

In Chapter 7, which is based on the paper *Ahmad, Buchardt, and Furrer (2022)*, we consider the problem of computing market values of bonus payments in multi-state with-profit life insurance. We consider the bonus scheme known as additional benefits, where dividends are used to buy extra benefits to the insured. The payment process in question takes the form

$$\mathrm{d}B(t) = \mathrm{d}B^\circ(t) + Q(t)\,\mathrm{d}B^\dagger(t),$$

where B° contains predetermined guaranteed payments and $B^\dagger$ contains unit bonus benefits, while $Q = \{Q(t)\}_{t\geq 0}$ is the process giving the number of units of additional benefits held. The focal point of the chapter is then to compute the market value of bonus payments, given by

$$V^b(0) = \mathbb{E}\left[\int_0^T e^{-\int_0^s r(v)\,\mathrm{d}v} Q(s)\,\mathrm{d}B^\dagger(s)\right],$$

where also the short rate $r = \{r(t)\}_{t\geq 0}$ is assumed to be stochastic of diffusion type, and where we have an underlying financial market $S = (S_0, S_1)$ consisting of a bank account S_0 (driven by r), and a risky asset S_1.

In full generality, $Q(t)$ will depend on the whole history on the financial market (S) and the states of the insured (Z) up to time t; it is $\mathcal{F}^S(t) \vee \mathcal{F}^Z(t)$–measurable.

This implies that the classic analytical methods encountered in Subsection 1.1.1 are not applicable for the computation of $V^b(0)$. We therefore take on a simulation approach, but where we only simulate financial risk and then use analytical methods for the outstanding insurance risk. Everything else being equal, this approach should be superior to a full blown Monte Carlo approach.

The key throughout the chapter is the so-called Q-modified transition probabilities, introduced as

$$p_{z_0 j}^Q(0,t) = \mathbb{E}\big[1_{(Z(t)=j)}Q(t) \,\big|\, \mathcal{F}^S(t)\big]. \tag{1.2.6}$$

From these, the market value of bonus can be calculated via simulations of financial scenarios using the representation, with $Z(0) \equiv z_0 \in \mathcal{J}$,

$$\begin{aligned} V^b(0) &= \mathbb{E}\left[\int_0^T e^{-\int_0^s r(v)\,\mathrm{d}v} A^b(0,\,\mathrm{d}s)\right], \\ A^b(0,\mathrm{d}s) &= \sum_{j\in\mathcal{J}} p_{z_0 j}^Q(0,s)\Big(b_j^\dagger(s) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}^\dagger(s)\mu_{jk}(s)\Big)\,\mathrm{d}s. \end{aligned}$$

Under suitable affinity assumptions on the dynamics of Q, we derive ordinary differential equations of Kolmogorov type for the Q-modified transition probabilities $p_{z_0 j}^Q(0,t)$. The type of differential equations obtained, as well as the very definition (1.2.6), bear a close resemblance to so-called retrospective reserves in single states considered in Norberg (1991, Subsection 5.B).

We then use the differential equation to formulate a numerical procedure for computation of $V^b(0)$. Throughout the chapter, we give special attention to a case where Q actually becomes $\mathcal{F}^S$-adapted. Here, $p_{z_0 j}^Q(0,t) = Q(t)p_{z_0 j}(0,t)$ and numerical procedures thus significantly simplifies into a direct computation of Q. In many cases, we think of this simplified procedure as an approximation to the more general procedure. We show a numerical example exactly illustrating this perspective, with the aim of comparing results between the two different procedures.

Although not mentioned in the chapter, the aforementioned relations between $p_{z_0 j}^Q$ and the retrospective reserves of Norberg (1991) further entails that it is possible to derive product integral representations akin to those of Bladt, Asmussen, and Steffensen (2020) for the Q-modified transition probabilities. This is also seen from the definition (1.2.6), which is closely related to the definition of the partial state-wise reserves (1.1.8). By a close inspection of the differential equation as well as the definitions, the representations should follow.

Chapter 2

Estimating absorption time distributions of general Markov jump processes

This chapter is based on the manuscript *Ahmad, Bladt, and Bladt (2022).*

Abstract

The estimation of absorption time distributions of Markov jump processes is an important task in various branches of statistics and applied probability. While the time-homogeneous case is classic, the time-inhomogeneous case has recently received increased attention due to its added flexibility and advances in computational power. However, commuting sub-intensity matrices are assumed, which in various cases limits the parsimonious properties of the resulting representation. This paper develops the theory required to solve the general case through maximum likelihood estimation, and in particular, using the expectation-maximization algorithm. A reduction to a piecewise constant intensity matrix function is proposed in order to provide succinct representations, where a parametric linear model binds the intensities together. Practical aspects are discussed and illustrated through the estimation of notoriously demanding theoretical distributions and real data, from the perspective of matrix analytic methods.

Keywords: Time-inhomogeneous Markov jump process; Inhomogeneous phase-type distribution; Parametric inference; EM algorithm; Poisson regression

2.1 Introduction

In this paper, we consider statistical estimation of distributions which are absorption times of general Markov jump processes, also known as inhomogeneous phase-type distributions (IPH). The data are the absorption times generated by independent samples of Markov jump processes until absorption, though the path is not observed. Thus, the incompleteness of the data is attended by an expectation-maximization (EM) algorithm, which allows for an effective maximum likelihood estimation. For practical purposes, we consider and implement the important special case where the underlying transition rates are piecewise constant.

Though time-inhomogeneous Markov jump processes have been classically used in many contexts, IPHs were only formally introduced in Albrecher and Bladt (2019) as the distribution of the absorption times in a time-inhomogeneous Markov jump process taking values on a finite state space where one state is absorbing and the remaining transient. They are a generalization of the classic phase-type distributions (PH), where the underlying Markov jump process is time-homogeneous (see, e.g., Bladt and Nielsen (2017) for an overview of the latter). These distributions may be used in situations where modeling tail behaviors different from the exponential, like e.g. heavy tails, is a concern, cf. the examples in Albrecher and Bladt (2019), where a subclass consisting of IPHs generated by intensity matrices which are given in terms of a single matrix scaled by some real non–negative function is considered. Within this subclass, the intensity matrices commute over time and thereby provide a link to the corresponding time-homogeneous PH distributions in terms of a parameter-dependent transformation. In this special case, the theory significantly simplifies and allows for more direct analysis. This is, for example, the case regarding statistical estimation, where Albrecher, Bladt, and Yslas (2022) develops an EM algorithm based on the parameter-dependent transformation so that the main engine basically uses the conventional EM algorithm known from PH fitting in Asmussen, Nerman, and Olsson (1996).

Since IPHs are absorption times of time-inhomogeneous Markov jump processes, they may naturally also be used for modeling processes that conceptually can be represented as evolving through states, e.g. in multi-state Markovian life insurance models (see, e.g., Hoem, 1969a; Norberg, 1991) where states (phases) relate to the different conditions of a policyholder in a time-dependent manner. This time-dependence would in general require non-commutative intensity matrices to provide meaningful models. Somewhat related, Albrecher et al. (2022) considers mortality modeling using IPHs, including age and time effects, though only the subclass of commuting matrices is examined here.

For time-inhomogeneous Markov jump processes, parametric modeling and maximum likelihood estimation of its transition rates based on the associated multivariate counting process is well-established in the literature; see, e.g., Andersen et al. (1993) for an overview. By assuming piecewise constant transition rates on a time grid (as an approximation), these methods are known to reduce to Poisson regressions based on aggregated occurrences and exposures in the different time intervals, cf., e.g., Aalen, Borgan, and Gjessing (2008, Section 5). This connection is particularly important in situations with aggregated data pooled into periodic intervals, like yearly observations. For example Poisson regression based on yearly observations is used in the Danish FSA's benchmark model for mortality risk, considered in Jarner and Møller (2015, Appendix 1), which is implemented in Danish life insurance and pension companies.

In this paper we extend the statistical fitting of IPHs from Albrecher, Bladt, and Yslas (2022) to the general class of IPHs, using these well-established techniques for parametric inference of time-inhomogeneous Markov jump processes as starting point; they constitute our (unobserved) complete data framework that generates the observations of IPHs and for which an EM algorithm is developed. This is in contrast to the approach in Albrecher, Bladt, and Yslas (2022), where the underlying homogeneous PH observations are seen as the building blocks. The general setting is, consequently, not reducible to the homogeneous case, and a non-trivial extension of the algorithm is required. In particular, the E-step is abstractly stated in terms of solutions of some differential equations, referred to as product integrals (see Gill and Johansen, 1990; Johansen, 1986), and the M-step involves numerical optimization.

Similarly to the completely observed data case, we identify the simplifications that arise in our EM algorithm from assuming piecewise constant transition rates on a time grid, whereby the E-step can be stated in terms of products of matrix exponentials to calculate a set of expected occurrences and exposures, and the M-step can be stated as performing maximum likelihood estimation in Poisson regressions akin to those of Aalen, Borgan, and Gjessing (2008, Section 5). This fully explicit algorithm allows for computational simplifications similar to those obtained in the complete data case and incurs increased computational performance while retaining flexibility. We also implement this algorithm and show some numerical examples of mortality modeling of Danish lifetimes as well as examples of fitting to theoretical distributions, confirming that the class of models does not suffer from some of the drawbacks that usual matrix analytic methods have. Another reason for allowing for different intensity matrices in different regions of the support is more pragmatic since it allows for fitting data that traditionally requires higher order IPHs. This could, e.g., be multi-modal data or skewed data. In such cases, we may obtain adequate fits in a discretized model of a much lower dimension.

One additional extension of our model appears during the M-step since the classic EM algorithm of Asmussen, Nerman, and Olsson (1996) has an explicit solution (number of jumps divided by total time spent in states; the so-called occurrence/exposure rates), while in our case we require parametrization of the transition rates to perform the required Poisson regressions. The canonical parametrization consisting of an intercept agrees with the simpler explicit solution. Fortunately, the added computational burden is low since standard software deals with generalized linear models in a stable and effective manner.

The remainder of the paper is structured as follows. In Section 2.2, we recall the inhomogeneous phase-type distribution (IPH). Then, in Section 2.3, we start out with an exposition of parametric inference of time-inhomogeneous Markov jump process, which will constitute the complete data case. Subsequently, we tackle the incomplete data problem and develop EM algorithms for general IPHs and those with piecewise constant transition rates. In Section 2.4, we consider an approach to a strong approximation of IPHs with piecewise constant transition using PH distributions, which may be useful for when a homogeneous representation is required. Section 2.5 is then devoted to numerical examples of our results. Finally, in Section 2.6, we present some possible extensions of our model, including a case where a pre-specified tail behavior is required.

2.2 Inhomogeneous phase-type distributions

Let $X = \{X(t)\}_{t\geq 0}$ be a time-inhomogeneous Markov jump process taking values on the finite state space $E = \{1, \dots, p, p+1\}$, $p \in \mathbb{N}$, where the states $\{1, \dots, p\}$ are transient and state $p+1$ is absorbing. Denote by $\boldsymbol{\alpha} = (\boldsymbol{\pi}, 0) = (\pi_1, \dots, \pi_p, 0)$ the initial distribution of X, and $\boldsymbol{\Lambda}(t) = \{\mu_{ij}(t)\}_{i,j\in E}$ the intensity matrix of X. The intensity matrix $\boldsymbol{\Lambda}(t)$ is then on the form

$$\boldsymbol{\Lambda}(t) = \begin{pmatrix} \boldsymbol{T}(t) & \boldsymbol{t}(t) \\ 0 & 0 \end{pmatrix},$$

where $\boldsymbol{T}(t)$ is the sub-intensity matrix function describing transitions between the transient states, and $\boldsymbol{t}(t) = -\boldsymbol{T}(t)\boldsymbol{e}$ consists of the transition rates to the absorbing state. Let τ denote the time until absorption of X, i.e.

$$\tau = \inf\{t \geq 0 \ : \ X(t) = p+1\}.$$

Following Albrecher and Bladt (2019), we then say that τ is inhomogeneous phase-type distributed (IPH) with representation $(\boldsymbol{\pi}, \boldsymbol{T}(\cdot))$, and we write $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T}(\cdot))$.

The transition probability matrix $\boldsymbol{P}(s,t) = \{p_{ij}(s,t)\}_{i,j\in E}$ of X, with elements

$$p_{ij}(s,t) = \mathbb{P}(\, X(t) = j \,|\, X(s) = i),$$

is given in terms of the product integral of the transition intensity matrix (see Gill and Johansen, 1990; Johansen, 1986):

$$\boldsymbol{P}(s,t) = \prod_s^t (\boldsymbol{I} + \boldsymbol{\Lambda}(u)\,\mathrm{d}u) = \begin{pmatrix} \bar{\boldsymbol{P}}(s,t) & \boldsymbol{e} - \bar{\boldsymbol{P}}(s,t)\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix},$$

where $\bar{\boldsymbol{P}}(s,t) = \{p_{ij}(s,t)\}_{i,j\in\{1,\dots,p\}}$ is the transition (sub-)probability matrix between the transient states,

$$\bar{\boldsymbol{P}}(s,t) = \prod_s^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u), \tag{2.2.1}$$

and $\boldsymbol{e} = (1,1,\dots,1)'$.

Together with the initial distribution $\boldsymbol{\pi}$, this gives the density and survival function of τ (see Theorem 2.2 in Albrecher and Bladt, 2019) as

$$f_\tau(x) = \boldsymbol{\pi}\bar{\boldsymbol{P}}(0,x)\boldsymbol{t}(x), \tag{2.2.2}$$

$$\bar{F}_\tau(x) = \boldsymbol{\pi}\bar{\boldsymbol{P}}(0,x)\boldsymbol{e}. \tag{2.2.3}$$

In this paper, we consider the statistical fitting of IPHs based on independent observations. Although in Albrecher, Bladt, and Yslas (2022) an expectation-maximization (EM) algorithm was devised for the case where $\boldsymbol{T}(t) = \lambda(t)\boldsymbol{T}$ (for parametric $\lambda(t)$ intensity functions), which implies that $\boldsymbol{T}(t)$ commute for different t, no statistical model where $\boldsymbol{T}(\cdot)$ are non-commutative has been considered in the literature. This is a drawback of significant concern for certain applications, which we seek to remedy in this paper as our main contribution; we provide a general EM algorithm and implement it in the case of a piecewise constant intensity matrix function.

2.2.1 IPHs with piecewise constant intensity matrices

We now consider a discretization of the time axis, where in each sub-interval, a different constant intensity matrix is defined. The purpose of this specification is two-fold. First, we seek to provide a statistical methodology for the non-commutative case, which will ease the fitting of heterogeneous data with lower matrix dimensions than previously considered. Second, and perhaps less obvious, is the generalization of discretized non-matrix versions of our model, which require a large number of intervals to provide a satisfactory approximation to the behavior of real data. In this context, the introduction of matrix parameters will allow for more flexible interpolation within sub-intervals, reducing the mesh size of the discretization.

Construct a grid $s_0 = 0 < s_1 < \cdots < s_{K-1} < \infty = s_K$, so that $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T}(\cdot))$, where

$$\boldsymbol{T}(s) = \boldsymbol{T}_k = \left\{\mu_{ij}^k\right\}_{i,j=1,\dots,p}, \quad s \in (s_{k-1}, s_k], \quad k = 1,\dots,K, \tag{2.2.4}$$

and introducing $k(x)$ as the unique $k \in \{1,\dots,K\}$ satisfying that $x \in (s_{k-1}, s_k]$, then the product integral formula (2.2.1) for the (sub-)probability matrix between the transient states reduces to a product of matrix exponentials:

$$\bar{\boldsymbol{P}}(s,t) = \mathrm{e}^{\boldsymbol{T}_{k(s)}\left(s_{k(s)}-s\right)} \left(\prod_{\ell=k(s)+1}^{k(t)-1} \mathrm{e}^{\boldsymbol{T}_{\ell}(s_{\ell}-s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{T}_{k(t)}\left(t-s_{k(t)-1}\right)},$$

with the convention that the empty product equals the identity matrix. The density (2.2.2) and survival function (2.2.3) then in particular reduces to:

$$\bar{F}_{\tau}(x) = \boldsymbol{\pi} \left(\prod_{\ell=1}^{k(x)-1} \mathrm{e}^{\boldsymbol{T}_{\ell}(s_{\ell}-s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{T}_{k(x)}\left(x-s_{k(x)-1}\right)} \boldsymbol{e},$$

$$f_{\tau}(x) = \boldsymbol{\pi} \left(\prod_{\ell=1}^{k(x)-1} \mathrm{e}^{\boldsymbol{T}_{\ell}(s_{\ell}-s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{T}_{k(x)}\left(x-s_{k(x)-1}\right)} \boldsymbol{t}_{k(x)}.$$

These expressions may be regarded as discrete approximations to their corresponding product integral expressions of the general case but have the advantage of being computationally much lighter to evaluate. Indeed, algorithms for computing the exponential of a matrix are varied and efficient, while product integration must be computed by numerically solving differential equations of increased complexity.

The density of τ may be discontinuous at the interval endpoints, which define the constant matrices. Indeed, consider e.g. $f_{\tau}(s_1-)$ and $f_{\tau}(s_1+)$. Since the matrix exponential is continuous, we have that

$$f_{\tau}(s_1-) = \lim_{\epsilon\downarrow 0} \boldsymbol{\pi}\mathrm{e}^{\boldsymbol{T}_1(s_1-\epsilon)}\boldsymbol{t}_0 = \boldsymbol{\pi}\mathrm{e}^{\boldsymbol{T}_0 s_1}\boldsymbol{t}_0$$

while

$$f_{\tau}(s_1+) = \lim_{\epsilon\downarrow 0} \boldsymbol{\pi}\mathrm{e}^{\boldsymbol{T}_0 s_1}\mathrm{e}^{\boldsymbol{T}_1\epsilon}\boldsymbol{t}_1 = \boldsymbol{\pi}\mathrm{e}^{\boldsymbol{T}_0 s_1}\boldsymbol{t}_1.$$

Hence $f_{\tau}(s_1-)$ and $f_{\tau}(s_1+)$ may differ if $\boldsymbol{t}_0 \neq \boldsymbol{t}_1$, and similarly for all the other grid points. On the other hand, if all $\boldsymbol{t}_k = \boldsymbol{t}$ then the density for τ is continuous. Similarly, a sufficient condition for differentiability at all points is that $-\boldsymbol{T}_k^2\boldsymbol{e}$ does not depend on k.

2.3 Estimation

This section introduces the main contribution of the paper, namely the maximum-likelihood estimation of general IPHs through the expectation-maximization (EM) algorithm, with a special emphasis on the case of piecewise constant transition rates.

We proceed sequentially: first, the completely observed case is reviewed; second, the incomplete data setting is built using the estimators from the previous case; finally, a simplified algorithm with piecewise constant transition rates is presented.

2.3.1 The complete data case

We now review some methods known from the inference of time-inhomogeneous Markov jump processes on finite state spaces based on complete observations of its trajectories. We refer to Andersen et al. (1993) for a detailed exposition on this.

Suppose that we observe $N \in \mathbb{N}$ i.i.d. realizations of the time-inhomogeneous Markov jump process X on some time interval $[0, T]$, where $T > 0$ is a given and fixed time horizon; represent the data by $\boldsymbol{X} = (X^{(1)}, \dots, X^{(N)})$. Denote by $\boldsymbol{N} = (N^{(1)}, \dots, N^{(N)})$ the corresponding data of the multivariate counting process, where $N^{(n)}$, $n = 1, \dots, N$, have components

$$N_{ij}^{(n)}(t) = \# \left\{ s \in (0, t] : X^{(n)}(s-) = i,\ X^{(n)}(s) = j \right\}.$$

Parametrizing the transition rates with a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is some finite-dimensional, parameter space with non-empty interior, such that,

$$\boldsymbol{T}(s) = \boldsymbol{T}(s; \boldsymbol{\theta}),$$

we have that the likelihood function for the joint parameter $(\boldsymbol{\pi}, \boldsymbol{\theta})$ is given by

$$\begin{aligned}
\mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}) &= \mathcal{L}_0^{\boldsymbol{X}}(\boldsymbol{\pi}) \prod_{\substack{i,j \in E \\ j \neq i}} \mathcal{L}_{ij}^{\boldsymbol{X}}(\boldsymbol{\theta}), \\
\mathcal{L}_0^{\boldsymbol{X}}(\boldsymbol{\pi}) &= \prod_{i=1}^{p} \pi_i^{B_i}, \qquad (2.3.1) \\
\mathcal{L}_{ij}^{\boldsymbol{X}}(\boldsymbol{\theta}) &= \prod_{n=1}^{N} \exp\left(\int_{(0,T]} \log(\mu_{ij}(s; \boldsymbol{\theta}))\, \mathrm{d}N_{ij}^{(n)}(s) - \int_0^T I_i^{(n)}(s) \mu_{ij}(s; \boldsymbol{\theta})\, \mathrm{d}s \right),
\end{aligned}$$

where, for $i \in E$ and $n \in \{1, \dots, N\}$,

$$I_i^{(n)}(s) = \mathbb{1}_{(X^{(n)}(s) = i)} \qquad \text{and} \qquad B_i = \sum_{n=1}^{N} I_i^{(n)}(0). \tag{2.3.2}$$

Here, $I_i^{(n)}(s)$ indicates if the n'th observation has a sojourn in state i at time s, and B_i denotes the total number of observations with initial state i; only the latter can be aggregated over observations due to the initial distribution not having a time-dependency.

The corresponding log-likelihood $L^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \log \mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta})$ then takes form

$$L^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}) = L_0^{\boldsymbol{X}}(\boldsymbol{\pi}) + \sum_{\substack{i,j \in E \\ j \neq i}} L_{ij}^{\boldsymbol{X}}(\boldsymbol{\theta}), \tag{2.3.3}$$

$$L_0^{\boldsymbol{X}}(\boldsymbol{\pi}) = \sum_{i=1}^{p} B_i \log(\pi_i), \tag{2.3.4}$$

$$L_{ij}^{\boldsymbol{X}}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(\int_{(0,T]} \log(\mu_{ij}(s;\boldsymbol{\theta}))\, \mathrm{d}N_{ij}^{(n)}(s) - \int_0^T I_i^{(n)}(s)\mu_{ij}(s;\boldsymbol{\theta})\, \mathrm{d}s \right), \quad (2.3.5)$$

from which we obtain the MLE of $(\boldsymbol{\pi}, \boldsymbol{\theta})$:

$$(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{\pi},\boldsymbol{\theta})} L^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}).$$

The product structure of the likelihood (2.3.1) (equivalently the additive structure of the log-likelihood) in $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ via $\mathcal{L}_0^{\boldsymbol{X}}$ respectively $\mathcal{L}_{ij}^{\boldsymbol{X}}$, $i, j \in E$, $j \neq i$, enables us to estimate these separately. Regarding $\boldsymbol{\pi}$, we may note (or confirm by direct calculation) that the likelihood $\mathcal{L}_0^{\boldsymbol{X}}$ is proportional to the likelihood obtained from viewing $(B_1, \ldots, B_p)$ as an observation from the Multinomial$(N, \boldsymbol{\pi})$-distribution, where N is considered fixed. This gives a closed-form expression for the MLE:

$$\hat{\pi}_i = \frac{B_i}{N}.$$

For $\boldsymbol{\theta}$, a closed form expression for the MLE is not available in general, and numerical methods for the optimization

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{\substack{i,j\in E\\ j\neq i}} L_{ij}^{\boldsymbol{X}}(\boldsymbol{\theta})$$

are required.

2.3.2 The complete data case with piecewise constant transition rates

We now assume that the transition rates $\mu_{ij}(\cdot;\boldsymbol{\theta})$ are piecewise constant on the form (2.2.4). The likelihood (2.3.1) then simplifies to

$$\mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{p} \pi_i^{B_i} \prod_{k=1}^{K} \prod_{\substack{i,j\in E\\ j\neq i}} \mu_{ij}^k(\boldsymbol{\theta})^{O_{ij}(k)} \exp\bigl(-E_i(k)\mu_{ij}^k(\boldsymbol{\theta})\bigr), \quad (2.3.6)$$

where $O_{ij}(k)$ is the total number of *occurrences* of transitions from state i to j in the time interval $(s_{k-1}, s_k]$, and $E_i(k)$ is the total time spent in state i in the time interval $(s_{k-1}, s_k]$, the so-called *local exposure*:

$$\begin{aligned} O_{ij}(k) &= \sum_{n=1}^{N} \int_{(s_{k-1},s_k]} \mathrm{d}N_{ij}^{(n)}(t), \\ E_i(k) &= \sum_{n=1}^{N} \int_{s_{k-1}}^{s_k} I_i^{(n)}(t)\, \mathrm{d}t. \end{aligned} \quad (2.3.7)$$

Remark 2.3.1. The likelihood (2.3.6) can be seen to reduce to the likelihood considered in Asmussen, Nerman, and Olsson (1996) by having $K = 1$ (corresponding to homogeneity) and no parametrization of the transition rates. △

Thus, in the case of piecewise constant transition rates, the occurrences and exposures in the different time intervals, along with the number of initiations in the different states,

$$\{(B_i,\, O_{ij}(k),\, E_i(k)), \quad k = 1, \ldots, K, \;\; i, j \in E, \; j \neq i\}$$

are sufficient statistics. In fact, the resulting likelihood (2.3.6) is proportional to the likelihood obtained from independent observations

$$\begin{aligned}
&(B_1, \ldots, B_p), \\
&(O_{ij}(k), \quad k = 1, \ldots, K, \;\; i, j \in E, \; j \neq i),
\end{aligned} \tag{2.3.8}$$

where

$$\begin{aligned}
(B_1, \ldots, B_p) \quad &\text{is} \quad \text{Multinomial}(N, \boldsymbol{\pi}) - \text{distributed}, \\
O_{ij}(k) \quad &\text{is} \quad \text{Poisson}\big(E_i(k)\mu_{ij}^k(\boldsymbol{\theta})\big) - \text{distributed},
\end{aligned} \tag{2.3.9}$$

with N and $E_i(k)$ considered fixed. Consequently, the MLE of $\boldsymbol{\pi}$ is (still) given by

$$\hat{\pi}_i = \frac{B_i}{N},$$

while the MLE of $\boldsymbol{\theta}$ is obtained from Poisson regressions of the occurrences against the different times on the grid, which can be carried out using standard software packages. For example, if $\mu_{ij}^k(\boldsymbol{\theta})$ is an exponential function in $\boldsymbol{\theta}$, a Poisson regression with log-link function and log-exposure as offsets can be carried out, corresponding to the fitting of the models:

$$\log(\mu_{ij}(s; \boldsymbol{\theta})) = \log(E_i) + \theta_{ij}^{(1)} + \theta_{ij}^{(2)} \cdot f^{(2)}(s), \tag{2.3.10}$$

for some suitable known function $f^{(2)}$, with a common choice being the identity. The predictions at s_k and at unit exposure are then the estimates of the transition rates, $\mu_{ij}^k(\hat{\boldsymbol{\theta}})$.

In the case where the parameters in $\boldsymbol{\theta}$ act as the (unknown) piecewise constant transition rates themselves, i.e. $\boldsymbol{\theta} = (\theta_{ij}^k)_{k=1,\ldots,K,\, i,j \in E,\, j \neq i}$ so that

$$\mu_{ij}^k(\boldsymbol{\theta}) = \theta_{ij}^k,$$

the MLE of $\boldsymbol{\theta}$ simplifies to so-called occurrence–exposure rates:

$$\hat{\theta}_{ij}^k = \frac{O_{ij}(k)}{E_i(k)}.$$

This is a special case where transition rates are estimated directly in a "non-parametric" way and can be retrieved by considering the s_k as a categorical (instead of numeric) variable in (2.3.10). The assumption of piecewise constant transition rates is often seen as an approximation to the general continuous versions obtained when the number of grid points tends to infinity. However, the resulting estimated models may be favorable even for coarser grid mesh sizes.

2.3.3 EM algorithm for IPHs

Suppose that we observe N i.i.d. realizations of IPHs with representation $(\boldsymbol{\pi}, \boldsymbol{T}(\cdot;\boldsymbol{\theta}))$ and represent the data by the vector $\boldsymbol{\tau} = (\tau^{(1)}, \ldots, \tau^{(N)})$. The data $\boldsymbol{\tau}$ is then considered as incomplete data of the whole Markov jump process X on $[0, T]$, where $T = \max_{n=1,\ldots,N} \tau^{(n)}$, and we employ an EM algorithm to estimate the parameter $(\boldsymbol{\pi}, \boldsymbol{\theta})$ based on the complete data likelihood considered in the previous subsections.

Let $\mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}$ denote the expectation under which the Markov jump process X has sub-intensity matrices $\boldsymbol{T}(\cdot;\boldsymbol{\theta})$ and initial distribution $\boldsymbol{\pi}$. The EM algorithm for estimation of $(\boldsymbol{\pi}, \boldsymbol{\theta})$ then consists of initializing with some value $(\boldsymbol{\pi}^{(0)}, \boldsymbol{\theta}^{(0)}) \in [0,1]^{p+1} \times \boldsymbol{\Theta}$, and then iteratively compute the conditional expected log-likelihood given the incomplete data $\boldsymbol{\tau}$ under the current parameter values $(\boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)})$, known as the E-step,

$$(\boldsymbol{\pi}, \boldsymbol{\theta}) \mapsto \bar{L}^{(m)}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})}\left[L^{\boldsymbol{X}}(\boldsymbol{\pi}, \boldsymbol{\theta}) \,\middle|\, \boldsymbol{\tau} \right], \quad m \in \mathbb{N}_0, \tag{2.3.11}$$

and then update the parameters to $(\boldsymbol{\pi}^{(m+1)}, \boldsymbol{\theta}^{(m+1)})$ by maximizing $\bar{L}^{(m)}$, known as the M-step. For notational convenience, we write, under some parameter $(\boldsymbol{\pi}, \boldsymbol{\theta})$,

$$\bar{\boldsymbol{P}}(s, t; \boldsymbol{\theta}) = \prod_{s}^{t} \big(\boldsymbol{I} + \boldsymbol{T}\big(u; \boldsymbol{\theta}\big)\mathrm{d}u\big) \tag{2.3.12}$$

for the transition (sub-)probability matrix in the transient states, and

$$\begin{aligned} f(x; \boldsymbol{\pi}, \boldsymbol{\theta}) &= \boldsymbol{\pi} \bar{\boldsymbol{P}}(0, x; \boldsymbol{\theta}) \boldsymbol{t}(x; \boldsymbol{\theta}), \\ \boldsymbol{t}(x; \boldsymbol{\theta}) &= -\boldsymbol{T}\big(x; \boldsymbol{\theta}\big)\boldsymbol{e}, \end{aligned} \tag{2.3.13}$$

for the corresponding density. To derive the conditional expected log-likelihood (2.3.11), we essentially need the distribution of the Markov jump process conditional on its absorption time. This is obtained in the following lemma.

Lemma 2.3.2. *Let $X = \{X(s)\}_{s\geq 0}$ be a time-inhomogeneous Markov jump process taking values on E with sub-intensity matrix function $\boldsymbol{T}(\cdot\,;\boldsymbol{\theta})$ and initial distribution $\boldsymbol{\pi}$. Let $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T}(\cdot\,;\boldsymbol{\theta}))$ be its corresponding absorption time. The conditional process*

$$Y(s) \overset{d}{=} X(s)\big|\tau, \quad s \in [0, \tau),$$

is then a time-inhomogeneous Markov jump process taking values on $\{1, \ldots, p\}$ with initial distribution

$$\widetilde{\pi}_i(\tau; \boldsymbol{\pi}, \boldsymbol{\theta}) = \frac{\pi_i \boldsymbol{e}_i' \bar{\boldsymbol{P}}(0, \tau; \boldsymbol{\theta}) \boldsymbol{t}(\tau; \boldsymbol{\theta})}{\boldsymbol{\pi} \bar{\boldsymbol{P}}(0, \tau; \boldsymbol{\theta}) \boldsymbol{t}(\tau; \boldsymbol{\theta})},$$

transition probabilities

$$\widetilde{p}_{ij}(t, s|\tau; \boldsymbol{\theta}) = \frac{\boldsymbol{e}_i' \bar{\boldsymbol{P}}(t, s; \boldsymbol{\theta}) \boldsymbol{e}_j \boldsymbol{e}_j' \bar{\boldsymbol{P}}(s, \tau; \boldsymbol{\theta}) \boldsymbol{t}(\tau; \boldsymbol{\theta})}{\boldsymbol{e}_i' \bar{\boldsymbol{P}}(t, \tau; \boldsymbol{\theta}) \boldsymbol{t}(\tau; \boldsymbol{\theta})},$$

and transition intensities

$$\widetilde{\mu}_{ij}(t|\tau;\boldsymbol{\theta}) = \mu_{ij}(t;\boldsymbol{\theta})\frac{\boldsymbol{e}_j'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}{\boldsymbol{e}_i'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}.$$

Proof. Let $j \in \{1,\dots,p\}$ and $t,s \geq 0$ such that $0 \leq t \leq s < \tau$ be given. Then it follows from the law of iterated expectations and the Markov property of X that, for $y > s$, we get the conditional survival probability

$$\begin{aligned}
\mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(s)=j)}\mathbb{1}_{(\tau>y)} \,\middle|\, \mathcal{F}^X(t)\right] &= \mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(s)=j)}\mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(\tau>y)} \,\middle|\, \mathcal{F}^X(s)\right] \,\middle|\, \mathcal{F}^X(t)\right] \\
&= \mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(s)=j)}\boldsymbol{e}_{X(s)}'\bar{\boldsymbol{P}}(s,y;\boldsymbol{\theta})\boldsymbol{e} \,\middle|\, \mathcal{F}^X(t)\right] \\
&= \boldsymbol{e}_{X(t)}'\bar{\boldsymbol{P}}(t,s;\boldsymbol{\theta})\boldsymbol{e}_j\boldsymbol{e}_j'\bar{\boldsymbol{P}}(s,y;\boldsymbol{\theta})\boldsymbol{e},
\end{aligned}$$

from which obtain the transition probabilities for Y:

$$\begin{aligned}
\mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(Y(s)=j)} \,\middle|\, \mathcal{F}^Y(t)\right] &= \mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(s)=j)} \,\middle|\, \mathcal{F}^X(t)\vee\sigma(\tau)\right] \\
&= \frac{-\frac{\partial}{\partial y}\left(\boldsymbol{e}_{X(t)}'\bar{\boldsymbol{P}}(t,s;\boldsymbol{\theta})\boldsymbol{e}_j\boldsymbol{e}_j'\bar{\boldsymbol{P}}(s,y;\boldsymbol{\theta})\boldsymbol{e}\right)\Big|_{y=\tau}}{f(\tau;\boldsymbol{e}_{X(t)}',\boldsymbol{\theta})} \\
&= \frac{\boldsymbol{e}_{X(t)}'\bar{\boldsymbol{P}}(t,s;\boldsymbol{\theta})\boldsymbol{e}_j\boldsymbol{e}_j'\bar{\boldsymbol{P}}(s,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}{\boldsymbol{e}_{X(t)}'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})} \\
&= \widetilde{p}_{X(t)j}(t,s|\tau;\boldsymbol{\theta}),
\end{aligned}$$

which by conditioning on $Y(t)=i$, $i\in\{1,\dots,p\}$, (which implies $X(t)=i$) yields the desired result. For the corresponding transition intensities, we get by definition of these,

$$\begin{aligned}
\widetilde{\mu}_{ij}(t|\tau;\boldsymbol{\theta}) &= \lim_{h\downarrow 0}\frac{\widetilde{p}_{ij}(t,t+h|\tau;\boldsymbol{\theta})}{h} \\
&= \frac{1}{\boldsymbol{e}_i'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}\lim_{h\downarrow 0}\frac{\boldsymbol{e}_i'\bar{\boldsymbol{P}}(t,t+h;\boldsymbol{\theta})\boldsymbol{e}_j}{h}\boldsymbol{e}_j'\bar{\boldsymbol{P}}(t+h,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta}) \\
&= \frac{1}{\boldsymbol{e}_i'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}\mu_{ij}(t;\boldsymbol{\theta})\boldsymbol{e}_j'\bar{\boldsymbol{P}}(t,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta}),
\end{aligned}$$

where we use the continuity of the transition (sub-)probability matrix (that is, continuity of product integrals) in the last equality. Finally, the initial distribution follows from similar techniques on

$$\mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(0)=i)}\mathbb{1}_{(\tau>y)}\right] = \mathbb{E}_{(\boldsymbol{\pi},\boldsymbol{\theta})}\left[\mathbb{1}_{(X(0)=i)}\boldsymbol{e}_{X(0)}'\bar{\boldsymbol{P}}(0,y)\boldsymbol{e}\right] = \pi_i\boldsymbol{e}_i'\bar{\boldsymbol{P}}(0,y)\boldsymbol{e}$$

which gives

$$\widetilde{\pi}_i(\tau;\boldsymbol{\pi},\boldsymbol{\theta}) = \mathbb{E}\big[\mathbb{1}_{(X(0)=i)}\,\big|\,\sigma(\tau)\big] = \frac{-\frac{\partial}{\partial y}\left(\pi_i \boldsymbol{e}_i'\bar{\boldsymbol{P}}(0,y)\boldsymbol{e}\right)\Big|_{y=\tau}}{f(\tau;\boldsymbol{\pi},\boldsymbol{\theta})} = \frac{\pi_i \boldsymbol{e}_i'\bar{\boldsymbol{P}}(0,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})}{\boldsymbol{\pi}\bar{\boldsymbol{P}}(0,\tau;\boldsymbol{\theta})\boldsymbol{t}(\tau;\boldsymbol{\theta})},$$

the desired initial distribution. □

Remark 2.3.3. In Hoem (1969b) and Norberg (1991), similar conditional distributions as those of Lemma 2.3.2 are derived. While they consider conditional distributions given future states, we consider conditional distributions given the time of absorption, which is a slight extension in which we include (particularly simple) future jump times in the conditioning. △

For $n \in \{1,\ldots,N\}$, $s \in (0,\tau^{(n)}]$ and $i,j \in E$, $j \neq i$, define the conditional expected statistics under the parameters $(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})$, $m \in \mathbb{N}_0$,

$$\begin{aligned}
\bar{B}_i^{(m)} &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}[B_i \,|\, \boldsymbol{\tau}], \\
\bar{I}_i^{(n,m)}(s) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}\Big[I_i^{(n)}(s)\,\Big|\,\boldsymbol{\tau}\Big], \\
\bar{N}_{ij}^{(n,m)}(s) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}\Big[N_{ij}^{(n)}(s)\,\Big|\,\boldsymbol{\tau}\Big].
\end{aligned} \tag{2.3.14}$$

We then obtain the conditional expected log-likelihood in the following result.

Theorem 2.3.4. *The conditional expected log-likelihood given the data $\boldsymbol{\tau}$ under the parameters $(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})$, $m \in \mathbb{N}_0$, is given by*

$$\bar{L}^{(m)}(\boldsymbol{\pi},\boldsymbol{\theta}) = \bar{L}_0^{(m)}(\boldsymbol{\pi}) + \sum_{\substack{i,j\in E\\ j\neq i}} \bar{L}_{ij}^{(m)}(\boldsymbol{\theta}),$$

$$\bar{L}_0^{(m)}(\boldsymbol{\pi}) = \sum_{i=1}^{p} \bar{B}_i^{(m)} \log(\pi_i),$$

$$\bar{L}_{ij}^{(m)}(\boldsymbol{\theta}) = \sum_{n=1}^{N}\left(\int_0^{\tau^{(n)}} \log(\mu_{ij}(s;\boldsymbol{\theta}))\,\mathrm{d}\bar{N}_{ij}^{(n,m)}(s) - \int_0^{\tau^{(n)}} \bar{I}_i^{(n,m)}(s)\mu_{ij}(s;\boldsymbol{\theta})\,\mathrm{d}s\right),$$

with all the non-zero conditional expected statistics given by, for $i,j \in \{1,\ldots,p\}$, $j \neq i$, and $s \in (0,\tau^{(n)}]$, $n \in \{1,\ldots,N\}$,

$$\bar{B}_i^{(m)} = \sum_{n=1}^{N} \frac{\pi_i^{(m)}\boldsymbol{e}_i'\bar{\boldsymbol{P}}\big(0,\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)\boldsymbol{t}\big(\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)}\big)},$$

$$\bar{I}_i^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}\big(0,s;\boldsymbol{\theta}^{(m)}\big)\,\boldsymbol{e}_i\boldsymbol{e}_i'\,\bar{\boldsymbol{P}}\big(s,\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)\boldsymbol{t}\big(\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)}\big)},$$

$$\mathrm{d}\bar{N}_{ij}^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}\big(0,s;\boldsymbol{\theta}^{(m)}\big)\,\boldsymbol{e}_i\mu_{ij}(s;\boldsymbol{\theta}^{(m)})\boldsymbol{e}_j'\,\bar{\boldsymbol{P}}\big(s,\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)\boldsymbol{t}\big(\tau^{(n)};\boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)}\big)}\,\mathrm{d}s,$$

and, for $j = p+1$,

$$\mathrm{d}\bar{N}_{i,p+1}^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}\big(0,s;\boldsymbol{\theta}^{(m)}\big)\,\boldsymbol{e}_i t_i\big(s;\boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)}\big)}\,\mathrm{d}\varepsilon_{\tau^{(n)}}(s),$$

where $\varepsilon_{\tau^{(n)}}$ is the Dirac measure in $\tau^{(n)}$.

Proof. It follows from the complete data log-likelihood (2.3.3)–(2.3.5), that the conditional expected log-likelihood (2.3.11) is given by

$$\bar{L}^{(m)}(\boldsymbol{\pi},\boldsymbol{\theta}) = \bar{L}_0^{(m)}(\boldsymbol{\pi}) + \sum_{\substack{i,j\in E\\ j\neq i}}^{p} \bar{L}_{ij}^{(m)}(\boldsymbol{\theta}),$$

where, for $i,j \in E$, $j \neq i$,

$$\bar{L}_0^{(m)}(\boldsymbol{\pi}) = \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}[\,L_0(\boldsymbol{\pi})\,|\,\boldsymbol{\tau}] = \sum_{i=1}^{p} \bar{B}_i^{(m)}\log(\pi_i),$$

$$\begin{aligned}\bar{L}_{ij}^{(m)}(\boldsymbol{\theta}) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}[\,L_{ij}(\boldsymbol{\theta})\,|\,\boldsymbol{\tau}] \\ &= \sum_{n=1}^{N}\bigg(\int_{(0,\tau^{(n)}]} \log(\mu_{ij}(s;\boldsymbol{\theta}))\,\mathrm{d}\bar{N}_{ij}^{(n,m)}(s) - \int_0^{\tau^{(n)}} \bar{I}_i^{(n,m)}(s)\mu_{ij}(s;\boldsymbol{\theta})\,\mathrm{d}s\bigg),\end{aligned} \tag{2.3.15}$$

where we have used Fubini's theorem in the last equality. To compute the conditional expectations appearing in (2.3.15), we get, by independence of the elements in $\boldsymbol{\tau}$ and Lemma 2.3.2, that for $i \in \{1,\dots,p\}$,

$$\bar{B}_i^{(m)} = \sum_{n=1}^{N} \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}\Big[\mathbb{1}_{(X^{(0)}(s)=i)}\,\big|\,\tau^{(n)}\Big] = \sum_{n=1}^{N} \widetilde{\pi}_i(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)}),$$

which by insertion provides the desired expression. For $\bar{I}_i^{(n,m)}$, we get

$$\begin{aligned}\bar{I}_i^{(n,m)}(s) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\,\boldsymbol{\theta}^{(m)})}\Big[\mathbb{1}_{(X^{(n)}(s)=i)}\,\big|\,\tau^{(n)}\Big] \\ &= \sum_{\ell=1}^{p} \widetilde{\pi}_\ell(\tau^{(n)};\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})\widetilde{p}_{\ell i}(0,s|\tau^{(n)};\boldsymbol{\theta}^{(m)}) \\ &= \sum_{\ell=1}^{p} \frac{\pi_\ell^{(m)}\boldsymbol{e}_\ell'\bar{\boldsymbol{P}}(0,s;\boldsymbol{\theta}^{(m)})\boldsymbol{e}_i\boldsymbol{e}_i'\bar{\boldsymbol{P}}(s,\tau^{(n)};\boldsymbol{\theta}^{(m)})\boldsymbol{t}(\tau^{(n)};\boldsymbol{\theta}^{(m)})}{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}(0,\tau^{(n)};\boldsymbol{\theta}^{(m)})\boldsymbol{t}(\tau^{(n)};\boldsymbol{\theta}^{(m)})} \\ &= \frac{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}(0,s;\boldsymbol{\theta}^{(m)})\boldsymbol{e}_i\boldsymbol{e}_i'\bar{\boldsymbol{P}}(s,\tau^{(n)};\boldsymbol{\theta}^{(m)})\boldsymbol{t}(\tau^{(n)};\boldsymbol{\theta}^{(m)})}{\boldsymbol{\pi}^{(m)}\bar{\boldsymbol{P}}(0,\tau^{(n)};\boldsymbol{\theta}^{(m)})\boldsymbol{t}(\tau^{(n)};\boldsymbol{\theta}^{(m)})}.\end{aligned}$$

For $\bar{N}_{ij}^{(n,m)}$, $j \in \{1, \ldots, p\}$, $j \neq i$, we proceed similarly, using the intensity process of $\{X^{(n)}(s)\}_{s<\tau^{(n)}}|\tau^{(n)}$ from Lemma 2.3.2, to get

$$\begin{aligned}
\bar{N}_{ij}^{(n,m)}(s) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\, \boldsymbol{\theta}^{(m)})}\left[\int_{(0,s]} \mathrm{d}N_{ij}^{(n)}(u) \,\middle|\, \tau^{(n)}\right] \\
&= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\, \boldsymbol{\theta}^{(m)})}\left[\int_0^s 1_{(X^{(n)}(u)=i)} \widetilde{\mu}_{ij}(u|\tau^{(n)}; \boldsymbol{\theta}^{(m)})\, \mathrm{d}u \,\middle|\, \tau^{(n)}\right] \\
&= \int_0^s \sum_{\ell=1}^{p} \widetilde{\pi}_\ell(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}) \widetilde{p}_{\ell i}(0, u|\tau^{(n)}; \boldsymbol{\theta}^{(m)}) \widetilde{\mu}_{ij}(u|\tau^{(n)}; \boldsymbol{\theta}^{(m)})\, \mathrm{d}u \\
&= \int_0^s \frac{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}(0, u; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_i \mu_{ij}(u; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_j' \bar{\boldsymbol{P}}(u, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}(\tau^{(n)}; \boldsymbol{\theta}^{(m)})}{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}(\tau^{(n)}; \boldsymbol{\theta}^{(m)})}\, \mathrm{d}u,
\end{aligned}$$

for which we take the dynamics in s to arrive at the desired result. Finally, for $j = p+1$, we may note that $N_{i,p+1}^{(n)}$ can be written as

$$N_{i,p+1}^{(n)}(s) = \mathbb{1}_{(s \geq \tau^{(n)})} \mathbb{1}_{(X^{(n)}(\tau^{(n)}-)=i)},$$

and so, using the same techniques as for the above quantities,

$$\begin{aligned}
&\bar{N}_{i,p+1}^{(n,m)}(s) \\
&= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})}\Big[\mathbb{1}_{(s \geq \tau^{(n)})} \mathbb{1}_{(X^{(n)}(\tau^{(n)}-)=i)} \,\Big|\, \tau^{(n)}\Big] \\
&= \mathbb{1}_{(s \geq \tau^{(n)})} \sum_{\ell=1}^{p} \widetilde{\pi}_\ell\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big) \widetilde{p}_{\ell i}\big(0, \tau^{(n)}|\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \\
&= \mathbb{1}_{(s \geq \tau^{(n)})} \sum_{\ell=1}^{p} \frac{\pi_\ell^{(m)} \boldsymbol{e}_\ell' \bar{\boldsymbol{P}}\big(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_i \boldsymbol{e}_i' \bar{\boldsymbol{P}}\big(\tau^{(n)}, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{t}\big(\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}(\tau^{(n)}; \boldsymbol{\theta}^{(m)})} \\
&= \mathbb{1}_{(s \geq \tau^{(n)})} \frac{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}\big(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_i t_i\big(\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}(\tau^{(n)}; \boldsymbol{\theta}^{(m)})},
\end{aligned}$$

where we use the continuity of product integrals in the second equality. Taking the dynamics in s now yields the desired result. □

The result shows that developing an EM algorithm for general IPH distributions significantly increases the computational complexity compared with the homogeneous case Asmussen, Nerman, and Olsson (1996) as well as the commuting inhomogeneous cases Albrecher, Bladt, and Yslas (2022). Indeed, since we no longer have a set of sufficient statistics for the different states and transitions, we must in the E-step compute the conditional expected log-likelihood $\bar{L}_{ij}^{(m)}$ directly. Evaluating this in a parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ involves a collection of product integral calculations, as opposed to matrix exponential calculations known from the two existing algorithms. Also, the subsequent M-step is no longer explicit with simple expressions, which is inherited from the fact that the complete data MLE is not explicit in general, and numerical optimization methods are therefore required to carry out the M-step.

As one may note from Subsection 2.2.1 and 2.3.2, the above mentioned computational complexities can be remedied by assuming piecewise constant transition rates on the form (2.2.4). We shall therefore assume this in the following to obtain our main algorithm and corresponding numerical examples; for completeness, we still provide the general EM algorithm in Appendix 2.A, since different simplifications may be drawn from the general case in the future.

Consider the complete data likelihood (2.3.6) in the case of piecewise constant transition rates, and recall the sufficient statistics (2.3.7) for the different states and transitions. Since the corresponding log-likelihood is linear in these sufficient statistics,

$$\log \mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi},\boldsymbol{\theta}) = \sum_{i=1}^{p} B_i \log(\pi_i) + \sum_{k=1}^{K} \sum_{\substack{i,j\in E\\ j\neq i}} \big(O_{ij}(k)\log(\mu_{ij}^k(\boldsymbol{\theta})) - E_i(k)\mu_{ij}^k(\boldsymbol{\theta})\big),$$

the E-step for the transitions simplifies so that it now suffices to compute the following conditional expected sufficient statistics, for $k = 1,\dots,K$,

$$\begin{aligned} \bar{B}_i^{(m)} &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})}[B_i \mid \boldsymbol{\tau}], \\ \bar{E}_i^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})}[E_i(k) \mid \boldsymbol{\tau}], \\ \bar{O}_{ij}^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\pi}^{(m)},\boldsymbol{\theta}^{(m)})}[O_{ij}(k) \mid \boldsymbol{\tau}], \end{aligned} \tag{2.3.16}$$

and then the M-step for updating $\boldsymbol{\theta}$ simplifies to the Poisson regression mentioned in Subsection 2.3.2, but where the occurrences and exposures are replaced by their conditional expectations computed in the E-step.

Based on Theorem 2.3.4 for the general cases, we immediately obtain these conditional expectations in Corollary 2.3.5 below. For notational convenience, we let $k^{(n)} = k(\tau^{(n)})$ denote the place on the grid that the n'th observation lies in, and, for $k_1, k_2 \in \{1,\dots,K\}$, $k_2 \geq k_1$, we define

$$\boldsymbol{A}(k_1,k_2;\boldsymbol{\theta}) = \prod_{\ell=k_1}^{k_2} \mathrm{e}^{\boldsymbol{T}_\ell(\boldsymbol{\theta})(s_\ell - s_{\ell-1})}. \tag{2.3.17}$$

Then the (sub-)probability matrix in the transient states (2.3.12) under some parameter $(\boldsymbol{\pi},\boldsymbol{\theta})$ as well as the corresponding density (2.3.13) can be written as

$$\begin{aligned} \bar{\boldsymbol{P}}(s,t;\boldsymbol{\theta}) &= \mathrm{e}^{\boldsymbol{T}_{k(s)}(\boldsymbol{\theta})\left(s_{k(s)}-s\right)} \boldsymbol{A}(k(s)+1,k(t)-1;\boldsymbol{\theta})\, \mathrm{e}^{\boldsymbol{T}_{k(t)}(\boldsymbol{\theta})\left(t-s_{k(t)-1}\right)}, \\ f(x;\boldsymbol{\pi},\boldsymbol{\theta}) &= \boldsymbol{\pi}\boldsymbol{A}(1,k(x)-1;\boldsymbol{\theta})\boldsymbol{t}_{k(x)}(\boldsymbol{\theta}). \end{aligned} \tag{2.3.18}$$

Corollary 2.3.5. *Suppose that the sub-intensity matrix function $\boldsymbol{T}$ is piecewise constant of the form* (2.2.4). *Then the conditional expected sufficient statistics* (2.3.16) *are given by, for $i, j \in \{1, \dots, p\}$, $j \neq i$,*

$$\bar{B}_i^{(m)} = \sum_{n=1}^{N} \frac{\pi_i^{(m)} \boldsymbol{e}_i' \bar{\boldsymbol{P}}(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}_{k^{(n)}}\big(\boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)}$$

$$\bar{E}_i^{(m)}(k) = \sum_{n=1}^{N} \frac{\int_{s_{k-1}\wedge\tau^{(n)}}^{s_k\wedge\tau^{(n)}} \boldsymbol{\pi} \bar{\boldsymbol{P}}(0, u; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_i \boldsymbol{e}_i' \bar{\boldsymbol{P}}(u, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}_{k^{(n)}}\big(\boldsymbol{\theta}^{(m)}\big) \mathrm{d}u}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)}$$

$$\bar{O}_{ij}^{(m)}(k) = \sum_{n=1}^{N} \frac{\int_{s_{k-1}\wedge\tau^{(n)}}^{s_k\wedge\tau^{(n)}} \boldsymbol{\pi} \bar{\boldsymbol{P}}(0, u; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_i \mu_{ij}^k(\boldsymbol{\theta}^{(m)}) \boldsymbol{e}_j' \bar{\boldsymbol{P}}(u, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{t}_{k^{(n)}}\big(\boldsymbol{\theta}^{(m)}\big) \mathrm{d}u}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)},$$

$$\bar{O}_{i,p+1}^{(m)}(k) = \sum_{n=1}^{N} \mathbb{1}_{(\tau^{(n)} \in (s_{k-1}, s_k])} \frac{\boldsymbol{\pi} \bar{\boldsymbol{P}}(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_i \boldsymbol{e}_i' \boldsymbol{t}_{k^{(n)}}(\boldsymbol{\theta}^{(m)})}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)},$$

with $\bar{\boldsymbol{P}}$ and f given as in (2.3.18).

Proof. By inserting the expressions for $O_{ij}(k)$ and $E_i(k)$ from (2.3.7) into (2.3.16) and using Theorem (2.3.4), we obtain the results for $\bar{O}_{ij}^{(m)}(k)$ and $\bar{E}_i^{(m)}(k)$. For $\bar{B}_i^{(m)}$, it follows from a direct application of Theorem (2.3.4). □

By writing out the exact expressions for $\bar{\boldsymbol{P}}$ and f given as in (2.3.18), we end up with Algorithm 2.1, which by Corollary 2.3.5 produces the required MLE estimation for IPHs with piecewise constant transition rates.

Remark 2.3.6. To compute the matrix $\boldsymbol{C}_k^{(n,m)}$, for fixed $n \in \{1, \dots, N\}$, $k \in \{1, \dots, K\}$, and $m \in \mathbb{N}_0$, this involves integrals of matrix exponentials, which may be computationally heavy. However, we can observe that by defining the block matrix

$$\boldsymbol{G}_k^{(n,m)} := \begin{pmatrix} \boldsymbol{T}_k(\boldsymbol{\theta}^{(m)}) & \boldsymbol{b}^{(n,m)}\big(k^{(n)} \wedge k + 1\big) \boldsymbol{a}_s^{(m)}\big(k^{(n)} \wedge k - 1\big) \\ \boldsymbol{0} & \boldsymbol{T}_k(\boldsymbol{\theta}^{(m)}) \end{pmatrix},$$

we obtain from Van Loan (1978) that

$$\mathrm{e}^{\boldsymbol{G}_k^{(n,m)}\left(\tau_{|k}^{(n)} - \tau_{|k-1}^{(n)}\right)} = \begin{pmatrix} \mathrm{e}^{\boldsymbol{T}_k\left(\boldsymbol{\theta}^{(m)}\right)\left(\tau_{|k}^{(n)} - \tau_{|k-1}^{(n)}\right)} & \boldsymbol{C}_k^{(n,m)} \\ \boldsymbol{0} & \mathrm{e}^{\boldsymbol{T}_k\left(\boldsymbol{\theta}^{(m)}\right)\left(\tau_{|k}^{(n)} - \tau_{|k-1}^{(n)}\right)} \end{pmatrix},$$

which reduces to a single matrix exponential calculation. Similar type of simplifications were noted in Albrecher, Bladt, and Yslas (2022, Remark 2). △

Algorithm 2.1 EM algorithm for IPHs with piecewise constant transition rates

Input: *Data points* $\boldsymbol{\tau} = (\tau^{(1)}, \dots, \tau^{(N)})$ *and initial parameters* $(\boldsymbol{\pi}^{(0)}, \boldsymbol{\theta}^{(0)})$.

0) Set $m := 0$.

1) *E-step:* Compute statistics for states $i, j \in \{1, \dots, p\}$, $j \neq i$, and grid points $k = 1, \dots, K$,

$$\bar{B}_i^{(m)} = \sum_{n=1}^{N} \frac{\pi_i^{(m)} b_i^{(n,m)}(1)}{\boldsymbol{\pi}^{(m)} \boldsymbol{b}^{(n,m)}(1)},$$

$$\bar{E}_i^{(m)}(k) = \sum_{n=1}^{N} \frac{\boldsymbol{e}_i' \boldsymbol{C}_k^{(n,m)} \boldsymbol{e}_i}{\boldsymbol{\pi}^{(m)} \boldsymbol{b}^{(n,m)}(1)},$$

$$\bar{O}_{ij}^{(m)}(k) = \sum_{n=1}^{N} \mu_{ij}^k\big(\boldsymbol{\theta}^{(m)}\big) \frac{\boldsymbol{e}_j' \boldsymbol{C}_k^{(n,m)} \boldsymbol{e}_i}{\boldsymbol{\pi}^{(m)} \boldsymbol{b}^{(n,m)}(1)},$$

$$\bar{O}_{i,p+1}^{(m)}(k) = \sum_{n=1}^{N} \mathbb{1}_{(k^{(n)}=k)} \frac{a_i^{(n,m)} \boldsymbol{e}_i' \boldsymbol{t}_k(\boldsymbol{\theta}^{(m)})}{\boldsymbol{\pi}^{(m)} \boldsymbol{b}^{(n,m)}(1)},$$

where

$$\boldsymbol{a}^{(n,m)} = \boldsymbol{a}_s^{(m)}(k^{(n)} - 1) e^{\boldsymbol{T}_{k^{(n)}}(\boldsymbol{\theta}^{(m)})\left(\tau^{(n)} - s_{k^{(n)}-1}\right)},$$

$$\boldsymbol{a}_s^{(m)}(\ell) = \boldsymbol{\pi}^{(m)} \boldsymbol{A}(1, \ell; \boldsymbol{\theta}^{(m)}),$$

$$\boldsymbol{b}^{(n,m)}(\ell) = \begin{cases} \boldsymbol{A}\big(\ell, k^{(n)} - 1; \boldsymbol{\theta}^{(m)}\big) e^{\boldsymbol{T}_{k^{(n)}}\left(\boldsymbol{\theta}^{(m)}\right)\left(\tau^{(n)} - s_{k^{(n)}-1}\right)} \boldsymbol{t}_{k^{(n)}}\big(\boldsymbol{\theta}^{(m)}\big) & \ell \leq k^{(n)} \\ \boldsymbol{t}_{k^{(n)}}\big(\boldsymbol{\theta}^{(m)}\big) & \ell > k^{(n)} \end{cases}$$

$$\boldsymbol{C}_k^{(n,m)} = \int_{\tau_{|k-1}^{(n)}}^{\tau_{|k}^{(n)}} \boldsymbol{c}_k^{(n,m)}(u) \,\mathrm{d}u,$$

$$\boldsymbol{c}_k^{(n,m)}(u) = e^{\boldsymbol{T}_k(\boldsymbol{\theta}^{(m)})\left(\tau_{|k}^{(n)} - u\right)} \boldsymbol{b}^{(n,m)}\big(k_{|n} + 1\big) \boldsymbol{a}_s^{(m)}\big(k_{|n} - 1\big) e^{\boldsymbol{T}_k(\boldsymbol{\theta}^{(m)})\left(u - \tau_{|k-1}^{(n)}\right)},$$

$$k_{|n} = k^{(n)} \wedge k,$$

$$\tau_{|k}^{(n)} = s_k \wedge \tau^{(n)}.$$

2) *M-step:* Update the parameters:

$$\hat{\pi}_i^{(m+1)} = \frac{\bar{B}_i^{(m)}}{N},$$

$\hat{\boldsymbol{\theta}}^{(m+1)}$: MLE of the regressions

$$\bar{O}_{ij}^{(m)}(k) \sim \mathrm{Pois}\big(\mu_{ij}^k(\boldsymbol{\theta}) \bar{E}_i^{(m)}(k)\big), \quad k = 1, \dots, K.$$

3) Set $m := m + 1$ and GOTO 1), until a stopping rule is satisfied.

Output: *Fitted parameters* $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})$.

2.4 An approximate homogeneous representation

In full generality, a phase-type approximation for any distribution is possible through the construction of Johnson and Taaffe (1988), where Erlang weights are constructed according to the increments of the target cumulative distribution function. However, when the target distribution arises as an absorption time of an inhomogeneous Markov jump process, recent developments in Bladt and Peralta (2022) provide an alternative pathwise approximation yielding strong approximants which are directly parametrized by the intensity matrix $\mathbf{\Lambda}$. Since phase-type distributions enjoy explicit formulas which their inhomogeneous counterparts may lack, such an approximation is practically relevant, and thus we outline it below. Section 2.5 presents some numerical examples of such an approximation.

Combining Theorem 4.2 and Proposition 4.3 in Bladt and Peralta (2022) yields, after some calculations, the following result:

Theorem 2.4.1 (Phase-type approximation)**.** *Let* $\tau \sim$ *IPH*$(\boldsymbol{\alpha}, \boldsymbol{T}(s))$ *where* $\boldsymbol{T}(s)$ *is given as in* (2.2.4)*. Define* $\boldsymbol{\alpha}^{(m)} = (\boldsymbol{\alpha}, \mathbf{0}, \dots, \mathbf{0})$*, and the* $mp \times mp$ *sub-intensity matrix*

$$\boldsymbol{T}^{(n,m)} = \begin{pmatrix} -n\boldsymbol{I} & n\boldsymbol{Q}_1^{(n)} & 0 & \dots & 0 \\ 0 & -n\boldsymbol{I} & n\boldsymbol{Q}_2^{(n)} & \dots & 0 \\ 0 & 0 & -n\boldsymbol{I} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -n\boldsymbol{I} \end{pmatrix}, \tag{2.4.1}$$

where, for $\ell = 1, \dots, m-1$*, and* $\omega_k(\ell, n) = E(s_k; \ell, n) - E(s_{k-1}; \ell, n)$ *(here,* $E(\cdot; a, b)$ *is the Erlang cdf with* a *stages and rate* b*),*

$$\boldsymbol{Q}_\ell^{(n)} = \sum_{k=1}^{K} \omega_k(\ell, n) \boldsymbol{T}_k / n + \boldsymbol{I}. \tag{2.4.2}$$

Then there exist $\tau^{(n,m)} \sim$ *PH*$(\boldsymbol{\alpha}^{(m)}, \boldsymbol{T}^{(n,m)})$ *such that*

$$\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}(|\tau - \tau^{(n,m)}| > \epsilon) = 0.$$

Moreover, the density of the resulting approximation reduces to

$$f_{\tau^{(n,m)}}(t) = \sum_{\ell=1}^{m-1} \left[\boldsymbol{\alpha}\boldsymbol{Q}_1^{(n)} \cdots \boldsymbol{Q}_{\ell-1}^{(n)}(\boldsymbol{I} - \boldsymbol{Q}_\ell^{(n)})\boldsymbol{e}\right] \frac{t^{\ell-1}}{(\ell-1)!} n^\ell \exp(-nt) \tag{2.4.3}$$

$$+ \left[\boldsymbol{\alpha}\boldsymbol{Q}_1^{(n)} \cdots \boldsymbol{Q}_{m-1}^{(n)}\boldsymbol{e}\right] \frac{t^{m-1}}{(m-1)!} n^m \exp(-nt). \tag{2.4.4}$$

Remark 2.4.2. The above approximation is very computationally efficient. Indeed, the $\boldsymbol{Q}_\ell^{(n)}$ only vary across ℓ and n through the scalar Erlang weights $\omega_k(\ell, n)$. In particular, fast calculation of the Erlang density weights is possible.

One implicit assumption which is relevant when applying the approximation is that n must be large enough to make $\boldsymbol{T}^{(n,m)}$ a proper sub-intensity matrix, which depends on the maximal absolute value of the diagonal elements of the $\boldsymbol{T}_k$ matrices. Additionally, the choice of m should be such that $m \geq n \cdot \max_{i=1,\ldots,N}\{\tau^i\}$. △

2.5 Numerical examples

This section presents some numerical illustrations of our above model on theoretical distributions as well as real data. In both cases, we require a straightforward extension of Algorithm 2.1 to when each data point has a weight associated with it. Practically speaking, this is straightforwardly dealt with by providing a weight in each contribution for the conditional expectations of the E-step and replacing N with the sum of weights in the E-step. This extension allows for the estimation of histograms, known distributions (considering a discrete version of the theoretical density), or more efficient calculations for when we have repeated values. We provide examples of the two latter uses. In all cases, we consider piecewise IPH distributions with continuous densities.

2.5.1 Fitting to a given distribution

It is well known that phase-type distributions struggle to fit peaked distributions where the peak does not happen close to the origin; that is, a large number of phases are required for adequate estimation. Thus, we first consider the estimation of the $\mathcal{N}(2,\, 1/2)$ theoretical distribution (left truncated at 0, as to have only positive values) by:

1. A piecewise IPH with large K and small p.
2. A PH approximation to the piecewise IPH fit, as per Theorem 2.4.1.
3. A small and large homogeneous PH, for comparison.

By "small" and "large," we have used subjective judgment, but we are somewhat limited by computational power for any dimensions far exceeding the ones presented here.

The idea is thus to use the density height as weights for a given grid (here, we take the mesh size to be $\Delta_t = 0.05$), which is used as the observations in Algorithm 2.4.1. Applying this procedure can be appreciated in the left panel of Figure 2.1. We see that a very small phase-type dimension ($p = 2$) is required to provide a good fit if we allow for piecewise constant rates at a small grid, in this case considering 41 sub-intervals on the interval $[0, 4]$. Since all matrices in each sub-intervals are intrinsically linked through Equation (2.3.10), the number of parameters is kept low. We also see how an effective phase-type approximation is possible using the construction of Theorem 2.4.1, providing a visually indistinguishable representation from the piecewise counterpart and which enjoys a pathwise convergence interpretation.

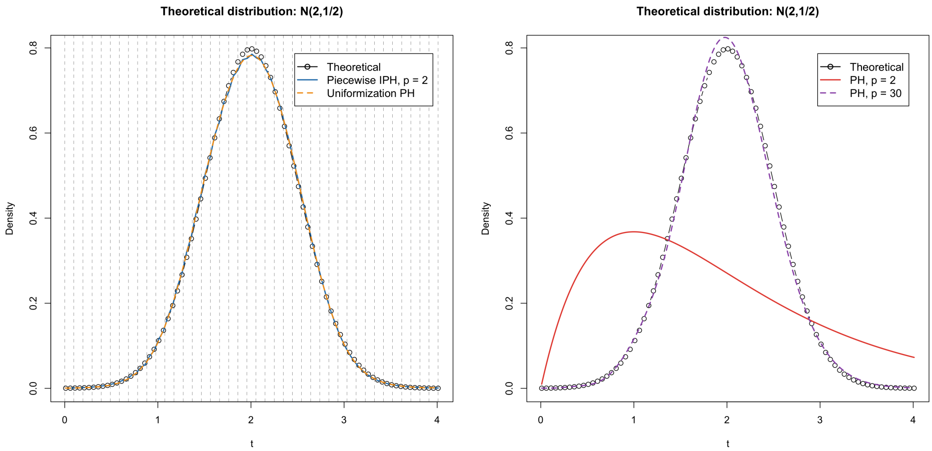

Figure 2.1: *Piecewise IPH (left) and PH (right) estimated densities to the theoretical* $\mathcal{N}(2, 1/2)$ *distribution.*

Note that the maximal absolute value of all diagonal matrices in each sub-interval for the piecewise IPH fit is 1056.8, which from the expression (2.4.1) implies that n should be at least above the latter value to obtain a proper phase-type sub-intensity matrix. We have thus chosen $n = 1500$, and then $m = n \cdot 4.01$, so the approximation is expected to be faithful up to the value 4.01. Thus the resulting phase-type approximation has state space dimension $p \times m = 12,030$, though the distribution is easy to manipulate, since formula (2.4.3) involves matrix calculus in terms of the original state space dimension p. In contrast, the right panel of Figure 2.1 shows that a 30-dimensional phase-type distribution cannot provide a similar quality of fit (let alone the 2-dimensional case). The EM algorithm which is required in this case (implemented as in Asmussen, Nerman, and Olsson (1996)) is comparatively slow for growing dimensions (and prohibitively slow for around $p = 50, 150$, depending on the language of implementation).

We now consider a more challenging setting with the aim of further showcasing the capabilities of our algorithm. Thus, we focus our attention on the mixture of $\mathcal{N}(2, 1/2)$ and $\mathcal{N}(4, 1/2)$ distributions, with a mixture weight of 0.55 (left truncated at 0, as to have only positive values), and we estimate two models:

1. A piecewise IPH with small K and medium p.

2. A homogeneous PH, for comparison.

For this multimodal density, we chose the breakpoints around valleys and summits of the theoretical density. An interesting comment is that choosing the breakpoints directly in the low point of a valley or exactly at the summit does not seems to be as effective. Given the chosen sub-intervals, we will use $p = 10$ since it seems to be the first dimension to capture both modes correctly. A PH approximation to the piecewise IPH fit, as per Theorem 2.4.1, is not possible in this setting since the estimated sub-intensity matrices for all sub-intervals have an overall largest absolute value in the diagonal equal to about $7.5 \cdot 10^{11}$, which implies that m is in the order of magnitude of 10^{12}, which is too large to make the computation of (2.4.3) feasible. As a general warning, we have found that for the most challenging density shapes, Theorem 2.4.1 will hold only theoretically, since practically it requires too many phases. This also confirms that sensible phase-type distributions do not suffice (including using the EM algorithm) in these cases.

The result of the estimation for this second case is provided in Figure 2.2, which shows the full strength of using piecewise IPH for heterogeneous data. We would like to comment that the dimension p and the number of subintervals K work together to provide an adequate fit and that a large K with small p does not work in this setting as it did for the previous unimodal distribution since the linear specification of $f^{(2)}$ in Equation (2.3.10) is no longer sufficient here. An alternative would be to consider spline specifications or higher polynomial terms. Here, we chose to increase the degrees of freedom by directly increasing p (in this case, to 10).

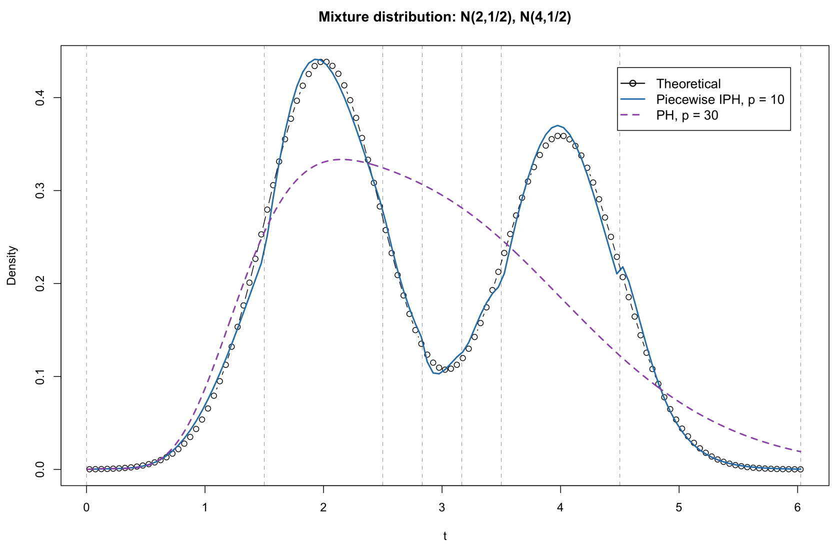

Figure 2.2: *Piecewise IPH and PH estimated densities to the theoretical mixture of* $\mathcal{N}(2, 1/2)$ *and* $\mathcal{N}(4, 1/2)$ *distributions, with mixing weight* 0.55.

Another feature that arises for estimated piecewise IPH distributions is the possible kink of the density at the endpoints of each sub-interval. These are not discontinuities and usually happen when the decreasing nature of a curve is not exponential, which is the case for gaussian decay. When examining the cumulative distribution function, the joining of the density of sub-intervals is differentiable; thus, the effect is not observable at that scale. These kinks also appear in the application to mortality modeling in the next section.

2.5.2 Mortality Modeling

The Human Mortality Database (`https://www.mortality.org/`) provides, among other things, mortality rates in a yearly resolution for several countries. We presently analyze the case of Danish males and females, from 2000 up to 2020. As before, we use as log-likelihood weights the implied density from the mortality rates (which is calculated as death to exposure ratio) and use the midpoints between ages as the observed ages (corresponding to the data $\boldsymbol{\tau}$). We divided for numerical purposes all data by 100 when estimating it. However, in the empirical versus fitted plotting, we have used the original scale (in any case, piecewise IPH are closed under scaling).

We have chosen the sub-intervals to provide more divisions for rapidly changing regions in the lifetime density, resulting in $K = 9$. We see from Figure 2.3 that, despite some possible kinks at the endpoints of intervals, the fit is remarkably well behaved, especially given the specific features that make modeling the entire lifetime distribution challenging: the sharp decrease after birth and the disruptions happening at around age 20 for both males and females. The increased mortality at the right endpoint also poses a challenge. The Gompertz-like behavior from around 30 to 100 is not in line with exponential decay; thus, regular sub-interval splits were required in this period. Finally, the resulting piecewise constant transition rates (in the log scale) are provided in Figure 2.4 for females and in Figure 2.5 for males, which are of interest for some disciplines that require mortality rate estimates, such as life insurance and pension applications.

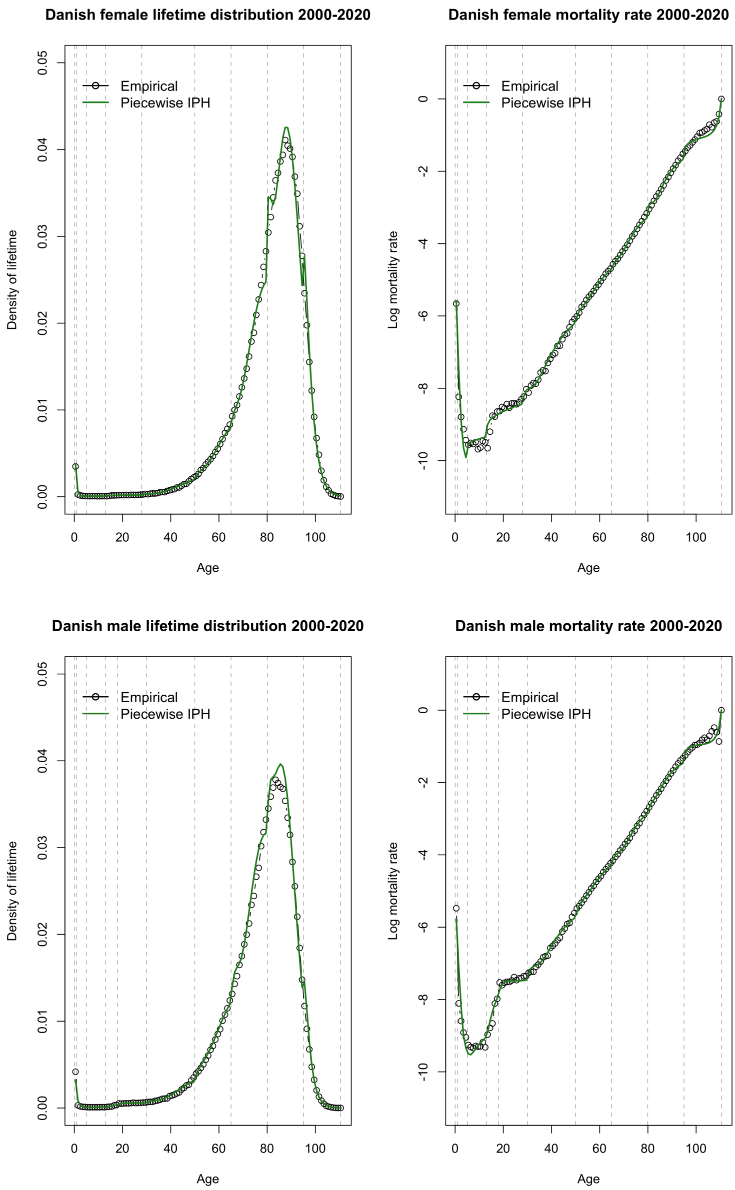

Figure 2.3: *Fitted versus empirical mortality curves using piecewise IPH distributions for Danish male and female populations from* 2000 *to* 2020*.*

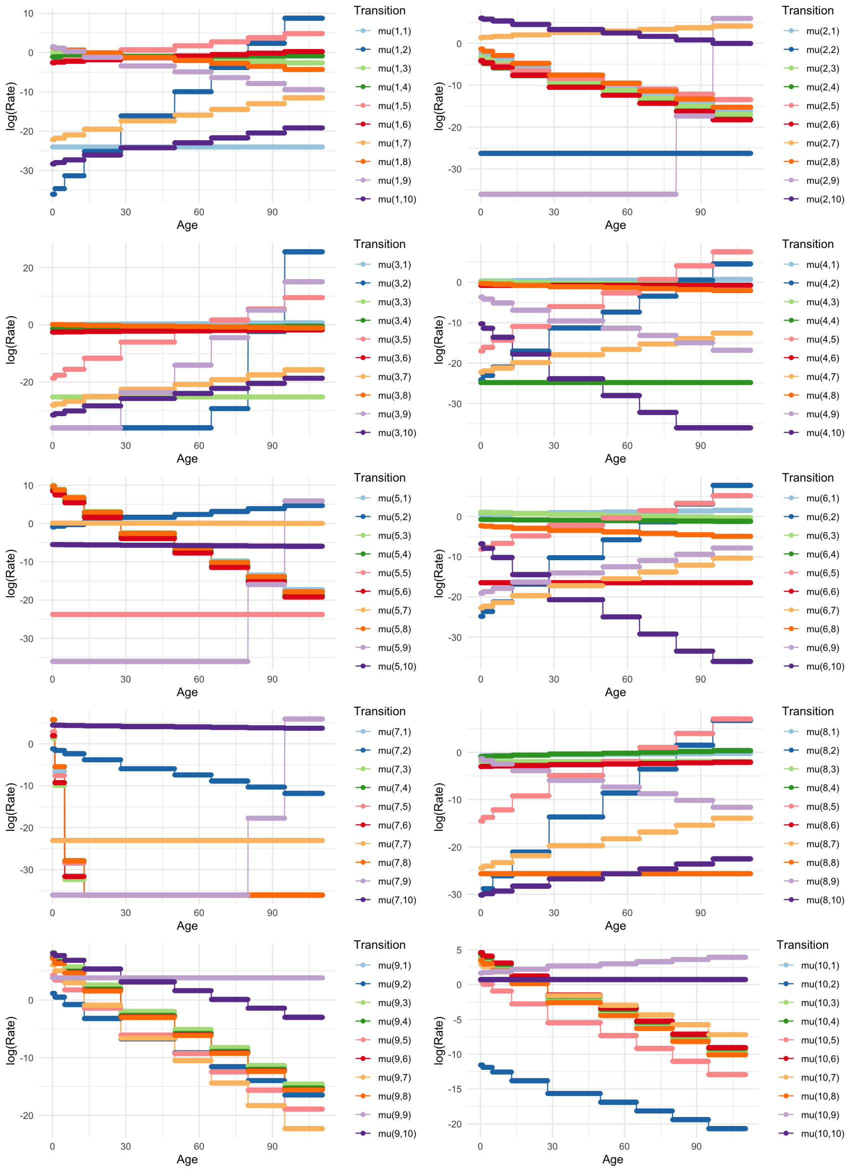

Figure 2.4: *Danish females: transition rates through time, for the fitted time-dependent sub-intensity matrix.*

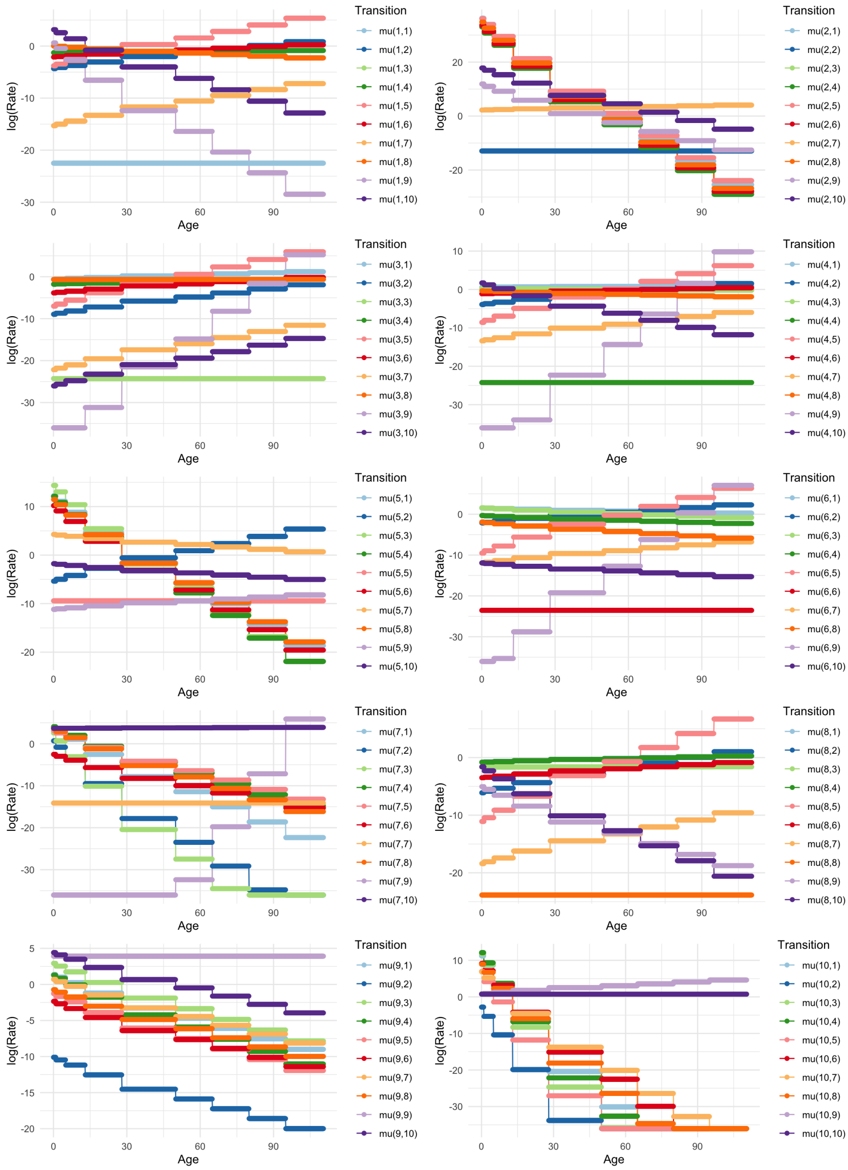

Figure 2.5: *Danish males: transition rates through time, for the fitted time-dependent sub-intensity matrix.*

2.6 Extensions

In this section, we discuss some possible extensions of theoretical and practical relevance that may be incorporated into our work but which is outside the scope of the present paper.

2.6.1 EM for IPHs with a pre-specified tail behavior

The focal point of the paper is to handle general IPHs with non-commutative sub-intensity matrix functions using piecewise constant transition rates as an approximation when the grid becomes finer. For a finite number of grid points, this construction implicitly implies an exponential tail behavior on the IPH distribution from the last grid point, which may not be suitable for applications on heavy-tailed data, e.g., non-life insurance data. However, it is straightforward to adapt our framework to an intrinsic possibility of obtaining a non-exponential tail behavior, using methods from Albrecher and Bladt (2019). The procedure goes as follows. Define a function λ by

$$\lambda(u) = \begin{cases} 1 & \text{if } u \le s_K \\ h(u) & \text{if } u > s_K, \end{cases}$$

for some non–negative function h, and a function g given in terms of its inverse by $g^{-1}(x) = \int_0^x \lambda(u)\mathrm{d}u$. Then $\widetilde{\tau} = g(\tau)$, where $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T}(\cdot))$ with $\boldsymbol{T}$ piecewise constant on the form (2.2.4), has a distribution with survival function

$$\bar{F}_{\widetilde{\tau}}(y) = \begin{cases} \boldsymbol{\pi} \left(\prod_{\ell=1}^{k(y)-1} \mathrm{e}^{\boldsymbol{T}_\ell (s_\ell - s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{T}_K (y - s_{K-1})} \boldsymbol{e} & \text{if } y \le s_{K-1} \\ \boldsymbol{\pi} \left(\prod_{\ell=1}^{K-1} \mathrm{e}^{\boldsymbol{T}_\ell (s_\ell - s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{T}_K \int_{s_{K-1}}^{y} h(u)\mathrm{d}u} \boldsymbol{e} & \text{if } y > s_{K-1}. \end{cases} \tag{2.6.1}$$

Hence, an extension of Algorithm 2.1 is possible for the model (2.6.1) with a pre-specified tail behavior according to the function h. Indeed, it suffices to apply the transformation $g^{-1}(x)$ of the data at the beginning of each step to reduce to the piecewise constant case, apply one EM step of Algorithm 2.1, and then optimize the parameter of the h function.

2.6.2 Censoring and truncation

In survival and event history analysis, one must take into account censoring and truncation mechanisms in the statistical estimation, see, e.g., Andersen et al. (1988) for a survey. This naturally also applies to the estimation of IPHs and PHs, as these are absorption times of Markov jump processes.

Incorporation of censoring mechanisms has long been established for estimation of PHs, cf. Olsson (1996), while the case of commuting matrices for IPHs is considered in Albrecher, Bladt, and Yslas (2022). As we have adapted Asmussen, Nerman, and Olsson (1996) to the inhomogeneous case by taking methods from Andersen et al. (1993) as onset, we believe that it is straightforward to incorporate the censoring mechanisms of Olsson (1996) to our model by adapting said paper to the inhomogeneous case taking methods from Andersen et al. (1988) as the onset.

To the best of our knowledge, the incorporation of truncation mechanisms has not yet been established for the estimation PHs or IPHs. We do not believe that this extension is straightforward in either framework, as one would need to consider conditional distributions of PHs and IPHs in developing the EM algorithm. These conditional distributions do not simplify to path-independent distributions as seen for fully observed Markov processes.

2.6.3 Covariate information

It is straightforward to include time-independent covariate information in our statistical model. Indeed, in the Poisson regressions in the EM algorithms presented for the piecewise-constant transition rate case, one may incorporate any (possibly transformed) covariate vector linearly, though each individual would have their own intensity matrices (which the other parts of the algorithm need to keep track of). The mortality modeling of Danish lifetimes in Subsection 2.5.2 is an example where sex could be used as a covariate.

Acknowledgments

Martin Bladt would like to acknowledge financial support from the Swiss National Science Foundation Project 200021_191984. The collaboration between the authors was strengthened by a research stay of Jamaal Ahmad at the University of Lausanne facilitated by Hansjörg Albrecher and partly supported by Oberstløjtnant Max Nørgaard og Hustru Magda Nørgaards Fond under File No. WZ 632-0055/21-2000.

2.A The general EM algorithm

Algorithm 2.2 EM algorithm for general IPHs

Input: Data points $\boldsymbol{\tau} = (\tau^{(1)}, \dots, \tau^{(N)})$ and initial parameters $(\boldsymbol{\pi}^{(0)}, \boldsymbol{\theta}^{(0)})$.

0) Set $m := 0$.

1) *E-step:* For $i \in \{1, \dots, p\}$, compute the conditional statistics for the initial state,

$$\bar{B}_i^{(m)} = \sum_{n=1}^{N} \frac{\pi_i^{(m)} \boldsymbol{e}_i' \bar{\boldsymbol{P}}\big(0, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{t}\big(\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)},$$

and, for $j \in E$, $j \neq i$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (on a suitable grid), compute the conditional expected log-likelihood for the transitions:

$$\bar{L}_{ij}^{(m)}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(\int_{(0, \tau^{(n)}]} \log(\mu_{ij}(s; \boldsymbol{\theta})) \mathrm{d}\bar{N}_{ij}^{(n,m)}(s) - \int_0^{\tau^{(n)}} \bar{I}_i^{(n,m)}(s) \mu_{ij}(s; \boldsymbol{\theta}) \mathrm{d}s \right),$$

where, for $j \neq p + 1$,

$$\bar{I}_i^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}\big(0, s; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_i \boldsymbol{e}_i' \bar{\boldsymbol{P}}\big(s, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{t}\big(\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)},$$

$$\mathrm{d}\bar{N}_{ij}^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}\big(0, s; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_i \mu_{ij}(s; \boldsymbol{\theta}^{(m)}) \boldsymbol{e}_j' \bar{\boldsymbol{P}}\big(s, \tau^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{t}\big(\tau^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)} \mathrm{d}s,$$

and, for $j = p + 1$,

$$\mathrm{d}\bar{N}_{i,p+1}^{(n,m)}(s) = \frac{\boldsymbol{\pi}^{(m)} \bar{\boldsymbol{P}}\big(0, s; \boldsymbol{\theta}^{(m)}\big)\, \boldsymbol{e}_i t_i\big(s; \boldsymbol{\theta}^{(m)}\big)}{f\big(\tau^{(n)}; \boldsymbol{\pi}^{(m)}, \boldsymbol{\theta}^{(m)}\big)} \,\mathrm{d}\varepsilon_{\tau^{(n)}}(s).$$

2) *M-step:* Update the parameters:

$$\hat{\pi}_i^{(m+1)} = \frac{\bar{B}_i^{(m)}}{N},$$

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\substack{i,j \in E \\ j \neq i}} \bar{L}_{ij}^{(m)}(\boldsymbol{\theta}).$$

3) Set $m := m + 1$ and GOTO 1) until a stopping rule is satisfied.

Output: *Fitted parameters* $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\theta}})$.

Chapter 3

Aggregate Markov models in life insurance: properties and valuation

This chapter is based on the manuscript *Ahmad, Bladt, and Furrer (2022).*

Abstract

In multi-state life insurance, an adequate balance between analytic tractability, computational efficiency, and statistical flexibility is of great importance. This might explain the popularity of Markov chain modeling, where matrix analytic methods allow for a comprehensive treatment. Unfortunately, Markov chain modeling is unable to capture duration effects, so this paper presents aggregate Markov models as an alternative. Aggregate Markov models retain most of the analytical tractability of Markov chains, yet are non-Markovian and thus more flexible. Based on an explicit characterization of the fundamental martingales, matrix representations of the expected accumulated cash flows and corresponding prospective reserves are derived for duration-dependent payments with and without incidental policyholder behavior. Throughout, special attention is given to a semi-Markovian case. Finally, the methods and results are illustrated in a numerical example.

Keywords: Multi-state modeling; semi-Markovianity; product integrals; expected cash flows; phase-type distributions

3.1 Introduction

In this paper, we propose a new class of multi-state models, the so-called aggregate Markov models, and study the valuation of life insurance liabilities for this class of models. The results established range from a characterization of the fundamental martingales to genuine computational schemes for the quantities of interest, including prospective reserves. In contrast, the companion paper Ahmad and Bladt (2022a) deals with the statistical aspects.

The classic approach to multi-state modeling consists of Markov chain modeling, where the process governing the state of the insured Z is taken to be a time-inhomogeneous Markov chain and the payments between the insured and the insurer are required to consist of deterministic sojourn and transition payments. This approach dates back to at least Hoem (1969a), but has recently been given new life via matrix analytic methods, see Bladt, Asmussen, and Steffensen (2020) and Ahmad (2022). Although Markov chain modeling is attractive due to its inherent simplicity, it suffers from a number of defects, not at least its inability to properly capture duration effects. In recent years, semi-Markov modeling has therefore gained considerable attention. In semi-Markov modeling, see Hoem (1972), Helwich (2008), Christiansen (2012), and Buchardt, Møller, and Schmidt (2015), only the joint process (Z, U), where U denotes the time spent in the current state, is assumed to be Markovian, and the sojourn and transition payments are allowed to depend on U. Unfortunately, semi-Markov modeling entails less analytical tractability and increased computational load.

In an aggregate Markov model, each observable state is assumed to consist of multiple unobservable sub-states, and only the full model consisting of all sub-states is assumed to be Markovian. Different to the classic Markov chains, which require the Markov property already for the observable states, aggregate Markov models are non-Markovian and thus flexible, yet they retain most of the analytical tractability of Markov chains. In particular, matrix analytic methods related to inhomogeneous phase-type distributions, confer with Albrecher and Bladt (2019), are applicable and lead to a unifying and transparent treatment.

Phase-type distributions have a long history of extensive use in applied probability. They have been employed in areas such as queueing theory Asmussen (2003), Neuts (1981, 1989), and Latouche and Ramaswami (1999), actuarial science Asmussen and Albrecher (2010) and Bladt and Nielsen (2017), and telecommunications Asmussen (2003) and Latouche and Ramaswami (1999), where the phase-type assumption leads to exact and in many cases explicit formulas for properties such as waiting time distributions, queue length, ruin probabilities, and buffer overflows.

Both homogeneous as well as inhomogeneous phase-type distributions are dense in the class of distributions on the positive reals, confer with Bladt and Nielsen (2017), and therefore able to approximate any non-negative distribution arbitrarily close – in the sense of weak convergence as the number of phases increases to infinity. Hence the class of phase-type distributions has been considered as striking a balance between tractability and generality. Inhomogeneous phase-type distributions may be used instead of homogeneous ones, and this might be particularly relevant if the tail behavior is known to be different from exponential, see Albrecher and Bladt (2019).

In the area of queuing theory, so-called quasi-birth-and-death (QBD) processes have been extensively studied, confer with Latouche and Ramaswami (1999), as a model for the number of customers in a queue. They constitute the time-homogeneous analogue to the aggregate Markov models considered here.

The first main contribution of the paper is an explicit characterization of the martingales for the associated counting processes N, which reveals that aggregate Markov models may be highly non-Markovian. For many practical purposes, less might suffice. To this end, we provide a sort of *reset property* under which the aggregate Markov model is actually semi-Markovian. The second main contribution of the paper are matrix representations for the expected accumulated cash flows, and hereby the prospective reserves, for duration-dependent payments with and without incidental policyholder behavior. Special attention is given to the case where the payments are duration independent; here our results indicate that aggregate Markov modeling may hold a competitive advantage over semi-Markov modeling.

The remainder of the paper is structured as follows. Section 3.2 provides some background, with Subsection 3.2.1 devoted to the basics of inhomogeneous phase-type distributions and Subsection 3.2.2 to the basics of multi-state modeling in life insurance. These subsections might be passed over by readers who are familiar with the subject matter. In Section 3.3, we introduce the aggregate Markov models and state the aforementioned reset property. The main contributions take place in Section 3.4 and Section 3.5. The former is devoted to the distributional properties of Z, including the characterization of the fundamental martingales, while the latter deals with the valuation of life insurance liabilities and contains, in particular, matrix representations of the expected accumulated cash flows. To showcase the practical potential of aggregate Markov modeling, Section 3.6 concludes with a numerical example. Proofs may be found in Appendix 3.A.

3.2 Preliminaries

Before introducing the setting of the paper, we provide some background. Subsection 3.2.1 contains a short review on inhomogeneous phase-type distributions, which play a critical role later. This review is followed by Subsection 3.2.2, which collects some insights on multi-state modeling in life insurance, in particular in relation to Markov chain and semi-Markov modeling, hereby motivating our aggregate setup. The actual presentation of our setup is postponed to Section 3.3.

In what follows and throughout the paper, we denote the product integral of a square matrix function $\boldsymbol{A}(x)$ as

$$\boldsymbol{F}(t,s) = \prod_t^s (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x),$$

where $\boldsymbol{I}$ is the identity matrix. Under suitable regularity conditions, it may equivalently be cast as the solution to Kolmogorov's forward and backward differential equations:

$$\begin{aligned}
\frac{\partial}{\partial s}\boldsymbol{F}(t,s) &= \boldsymbol{F}(t,s)\boldsymbol{A}(s), & \boldsymbol{F}(t,t) &= \boldsymbol{I}, \\
\frac{\partial}{\partial t}\boldsymbol{F}(t,s) &= -\boldsymbol{A}(t)\boldsymbol{F}(t,s), & \boldsymbol{F}(s,s) &= \boldsymbol{I}.
\end{aligned}$$

For a survey on product integration, we refer to Gill and Johansen (1990) and, concerning applications to life insurance, also Milbrodt and Stracke (1997) and Bladt, Asmussen, and Steffensen (2020).

3.2.1 Inhomogeneous phase-type distributions

In this subsection, we review the notion of inhomogeneous phase-type (IPH) distributions introduced in Albrecher and Bladt (2019). Consider a smooth and suitably regular time-inhomogeneous Markov jump process $X = \{X(t)\}_{t\geq 0}$ on the finite state space $\mathcal{J} = \{1,\dots,J-1,J\}$, where the states $\{1,\dots,J-1\}$ are transient while J is absorbing. The transition intensity matrix function $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j\in\mathcal{J}}$ of X is then on the form

$$\boldsymbol{M}(t) = \begin{pmatrix} \boldsymbol{T}(t) & \boldsymbol{t}(t) \\ 0 & 0 \end{pmatrix},$$

where $\boldsymbol{T}(t)$ is a sub-intensity matrix function consisting of transition rates between the transient states and $\boldsymbol{t}(t) = -\boldsymbol{T}(t)\boldsymbol{1}_J$ is a column vector of transition rates to the absorbing state, the so-called exit rate vector function. Further, assume that $\mathbb{P}(X(0) = J) = 0$ and denote by $\boldsymbol{\pi} = (\pi_1,\dots,\pi_{J-1})$ the remaining vector of initial probabilities $\pi_j = \mathbb{P}(X(0) = j)$. The time until absorption, given by

$$\tau = \inf\{t \geq 0 \,:\, X(t) = J\},$$

is then said to be an inhomogeneous phase-type distribution with representation $(\boldsymbol{\pi},\boldsymbol{T})$, and we write $\tau \sim \mathrm{IPH}(\boldsymbol{\pi},\boldsymbol{T})$.

The transition probability matrix function $\boldsymbol{P}(t,s) = \{p_{ij}(t,s)\}_{i,j\in\mathcal{J}}$ with elements

$$p_{ij}(t,s) = \mathbb{P}(X(s) = j \mid X(t) = i)$$

is given as the product integral of the transition intensity matrix function:

$$\boldsymbol{P}(t,s) = \prod_t^s (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x).$$

The probability density function $f(x)$ and distribution function $F(x)$ of τ may then be obtained through product integrals of the sub-intensity matrix function $\boldsymbol{T}(t)$:

$$f(x) = \boldsymbol{\pi} \prod_0^x (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{t}(x),$$
$$F(x) = 1 - \boldsymbol{\pi} \prod_0^x (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{1}_J.$$

From these, one finds the following conditional distribution:

$$\mathbb{P}(\tau > s+t \mid \tau > s) = \frac{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)}{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J} \prod_s^t (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J, \tag{3.2.1}$$

which entails that

$$\tau - s \mid \tau > s \sim \mathrm{IPH}(\boldsymbol{\alpha}(s), \boldsymbol{T}(s + \bullet)),$$

where $\boldsymbol{\alpha}(s)$ is given by

$$\boldsymbol{\alpha}(s) = \frac{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)}{\boldsymbol{\pi} \prod_0^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x)\boldsymbol{1}_J}.$$

In other words, for IPH distributions, the overshoot is again IPH-distributed.

Example 3.2.1. In the case of a single phase, that is $J = 2$, we have

$$\prod_t^s (\boldsymbol{I} + \boldsymbol{T}(x)\,\mathrm{d}x) = e^{-\int_t^s \mu_{12}(x)\,\mathrm{d}x},$$

giving the density and distribution functions

$$f(x) = e^{-\int_0^x \mu_{12}(v)\,\mathrm{d}v}\mu_{12}(x),$$
$$F(x) = 1 - e^{-\int_0^x \mu_{12}(v)\,\mathrm{d}v},$$

while the conditional distribution (3.2.1) takes the form

$$\mathbb{P}(\tau > s+t \mid \tau > s) = \frac{e^{-\int_0^s \mu_{12}(v)\,\mathrm{d}v}}{e^{-\int_0^s \mu_{12}(v)\,\mathrm{d}v}} e^{-\int_s^t \mu_{12}(v)\,\mathrm{d}v} = e^{-\int_s^t \mu_{12}(v)\,\mathrm{d}v}. \qquad \circ$$

3.2.2 Multi-state modeling

Insurance contracts may be modeled as a stream of payments $B = \{B(t)\}_{t\geq 0}$, benefits less premiums, between the insured and the insurer. In life insurance, including health and disability insurance and pensions, the payments depend on the state of the insured, leading to so-called multi-state modeling. In general, the state of the insured $Z = \{Z(t)\}_{t\geq 0}$ is a non-explosive jump process on a typically finite state space $\mathcal{J} = \{1, 2, \ldots, J\}$, $J \in \mathbb{N}$, while the payments are typically finite variation processes adapted to the information generated by Z.

Markov chain models

The most classic approach to multi-state modeling is (smooth) Markov chain models, where Z is taken to be a time-inhomogeneous Markov jump process (Markov chain) with suitably regular transition rates $\nu_{jk}(t)$, so that

$$\nu_{jk}(t) = \lim_{h\downarrow 0} \frac{\mathbb{P}(Z(t+h) = k \mid Z(t) = j)}{h}.$$

Using the standard convention $\nu_{jj}(t) = -\sum_{\substack{k\in\mathcal{J}\\ k\neq j}} \nu_{jk}(t)$, the square matrix function with indices $\nu_{jk}(t)$ is then the transition intensity matrix function of Z. In addition to the Markov assumption, the payments are assumed to take the form

$$\mathrm{d}B(t) = \sum_{j\in\mathcal{J}} \left(1_{(Z(t)=j)} b_j(t)\,\mathrm{d}t + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(t)\,\mathrm{d}N_{jk}(t) \right), \quad B(0) \in \mathbb{R},$$

for suitably regular deterministic sojourn payment rates $b_j(t)$ and transition payments $b_{jk}(t)$ depending only on time. Here N is the multivariate counting process associated to Z with components $N_{jk} = \{N_{jk}(t)\}_{t\geq 0}$ given by

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

Markov chain modeling dates back to at least Hoem (1969a) and was popularized in Norberg (1991).

Regarding the valuation of life insurance liabilities, calculating the so-called expected accumulated cash flow $A(t, s)$ is key. For Markov chain models, the expected accumulated cash flow $A(t, s)$ is given by $A(t, s) = \sum_{i\in\mathcal{J}} 1_{(Z(t)=i)} A_i(t, s)$, where

$$\begin{aligned} A_i(t, s) &= \mathbb{E}[B(s) - B(t) \mid Z(t) = i] \\ &= \sum_{j\in\mathcal{J}} \int_t^s \mathbb{P}(Z(u) = j \mid Z(t) = i) \Big(b_j(u) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} \nu_{jk}(u) b_{jk}(u) \Big) \mathrm{d}u. \end{aligned}$$

The transition probabilities, considered as a square matrix function, are given as the product integral of the transition rates, also considered as a square matrix function. In other words, the transition probabilities may be found simply by solving Kolmogorov's forward differential equations.

It is, of course, possible, and of interest, to relax both the Markov assumption as well as the structure of the payments. In doing so, it is critical to strike an adequate balance between analytic tractability, computational efficiency, and statistical flexibility.

Semi-Markov models

A more modern approach is semi-Markov modeling, see for instance Helwich (2008), Christiansen (2012), and Buchardt, Møller, and Schmidt (2015). There are two major differences between (smooth) semi-Markov modeling and (smooth) Markov chain modeling. First, the jump process Z describing the state of the insured is no longer required to be Markovian; rather, (Z, U) is assumed Markovian, where $U = \{U(t)\}_{t \geq 0}$ is the duration since the last transition given by

$$U(t) = \sup\big\{s \in [0,t] : Z(u) = Z(t) \text{ for all } u \in [t-s, t]\big\}. \tag{3.2.2}$$

Therefore, the model can no longer be described by transition rates that solely depend on time. Instead, the transition rates are now functions of both time and duration, written $\nu_{jk}(t,u)$.

Second, the payments take the more general form

$$\begin{aligned} \mathrm{d}B(t) &= \sum_{j \in \mathcal{J}} \bigg(1_{(Z(t)=j)} b_j(t, U(t)) \,\mathrm{d}t + \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} b_{jk}(t, U(t-)) \,\mathrm{d}N_{jk}(t) \bigg), \\ B(0) &\in \mathbb{R}, \end{aligned} \tag{3.2.3}$$

for suitably regular deterministic sojourn payment rates $b_j(t,u)$ and transition payments $b_{jk}(t,u)$ depending on time and duration.

For semi-Markov models, the expected accumulated cash flow $A(t,s)$ depends on both the current state and current duration. To clarify, it may actually be decomposed according to $A(t,s) = \sum_{i \in \mathcal{J}} 1_{(Z(t)=i)} A_{i,U(t)}(t,s)$, where

$$\begin{aligned} A_{i,u}(t,s) &= \mathbb{E}[B(s) - B(t) \,|\, Z(t) = i, U(t) = u] \\ &= \sum_{j \in \mathcal{J}} \int_t^s \int_0^{u+v-t} \bigg(b_j(v,z) + \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} \nu_{jk}(v,z) b_{jk}(v,z) \bigg) p_{ij}(t,u,v,\mathrm{d}z) \mathrm{d}v, \end{aligned}$$

$$p_{ij}(t,u,s,z) = \mathbb{P}(Z(s) = j, U(s) \leq z \,|\, Z(t) = i, U(t) = u).$$

The transition probabilities may be calculated by solving a system of integro-differential equations, confer with Buchardt, Møller, and Schmidt (2015, Section 3). Numerical methods for integro-differential equations can, generally speaking, be rather intricate. The implementation of semi-Markov models is, therefore, non-trivial and may carry some operational risk.

Aggregate Markov models

In this paper, we introduce a class of aggregate models that, similar to semi-Markov models, allow for added flexibility, such as duration dependence, but avoid some of the aforementioned numerical challenges posed by semi-Markov modeling.

Denote by $(T_n)_{n\in\mathbb{N}_0}$ the jump times of Z, where we employ the convention $T_0 = 0$. Returning to the case where Z is Markovian with transition rates $\nu_{jk}(t)$, recall that

$$\mathbb{P}(T_{n+1} > t \,|\, T_0, Z(T_0), T_1, Z(T_1), \ldots, T_n, Z(T_n) = j) = e^{\int_{T_n}^{t} \nu_{jj}(x)\,\mathrm{d}x}, \quad t \geq T_n.$$

We conclude that

$$T_{n+1} - T_n \,\Big|\, \big(T_i, Z(T_i)\big)_{i=0}^{n} \sim \mathrm{IPH}\big(1, \nu_{Z(T_n)Z(T_n)}(T_n + \bullet)\big).$$

In other words, the sojourn times follow one-dimensional IPH distributions that are mostly independent of the past history of the jump process. This paper considers instead jump processes with sojourn times admitting conditional IPH distributions of general dimension. Hereby we shall be able to capture, for instance, duration dependence while avoiding the need for intricate numerical methods.

3.3 Setup

In this section, we present the general setup of the paper. Subsection 3.3.1 introduces the probabilistic model for the state of the insured, while Subsection 3.3.2 introduces the payments between the insured and the insurer.

3.3.1 Probabilistic model

Similar to Subsection 3.2.2, let Z be a jump process governing the state of the insured, thus taking values in the finite set of (macrostates) $\mathcal{J} = \{1, 2, ..., J\}$, $J \in \mathbb{N}$. This set consists of biometric or behavioral states that are actually observed, for example active, disabled, free-policy, and dead. To allow for added flexibility, to each macrostate we may introduce additional sub-states (microstates) that are not observable.

To be specific, to each macrostate j, a number $d_j \geq 1$ of microstates are assigned. The resulting state space is therefore

$$E = \{\boldsymbol{j} = (j, \widetilde{j}) : j \in \mathcal{J}, \widetilde{j} \in \{1, 2, ..., d_j\}\},$$

and the total number of microstates is $\bar{d} = \sum_{j\in\mathcal{J}} d_j$. Elements of E are generally denoted by bold letters such as $\boldsymbol{j} \in E$. Now introduce a time-inhomogeneous Markov jump process $\boldsymbol{X} = \{\boldsymbol{X}(t)\}_{t\geq 0} = \{(X_1(t), X_2(t))\}_{t\geq 0}$ on the state space E with transition intensity matrix function $\boldsymbol{M}(t)$. Then $X_1(t)$ keeps track of the macrostate, that is $Z(t) = X_1(t)$, while $X_2(t)$ identifies the current microstate contingent on the state of $X_1(t)$.

The transition intensity matrix function $\boldsymbol{M}(t)$ can be written on the following block form:

$$\boldsymbol{M}(t) = \begin{pmatrix} \boldsymbol{M}_{11}(t) & \boldsymbol{M}_{12}(t) & \cdots & \boldsymbol{M}_{1J}(t) \\ \boldsymbol{M}_{21}(t) & \boldsymbol{M}_{22}(t) & \cdots & \boldsymbol{M}_{2J}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{M}_{J1}(t) & \boldsymbol{M}_{J2}(t) & \cdots & \boldsymbol{M}_{JJ}(t) \end{pmatrix}, \tag{3.3.1}$$

where $\boldsymbol{M}_{jj}(t)$ are sub-intensity matrix functions of dimension $d_j \times d_j$ providing transition rates between the microstates of macrostate j, and $\boldsymbol{M}_{jk}(t)$ are non-negative matrix functions of dimension $d_j \times d_k$ providing transition rates from microstates within macrostate j to microstates within macrostate k.

We denote an element of $\boldsymbol{M}(t)$ by $\mu_{\boldsymbol{j}\boldsymbol{k}}(t)$, $\boldsymbol{j}, \boldsymbol{k} \in E$. The off-diagonal elements are non-negative, providing the jump rates between different states, while the diagonal equals the negative of the row sums of the off-diagonal elements. Consequently, rows all sum to zero, so $\boldsymbol{M}(t)$ is a proper transition intensity matrix function.

For simplicity, we assume that $Z(0) = X_1(0) \equiv 1$. For a full model identification, it then suffices to specify the initial distribution $\boldsymbol{\pi}_1(0)$ of $X_2(0)$ among the microstates $1, 2, \ldots, d_1$. In other words, denoting the initial distribution of $\boldsymbol{X}$ by $\boldsymbol{\pi}$, we have that

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_1(0), \boldsymbol{0}).$$

The column vector function

$$\boldsymbol{m}_j(t) = -\boldsymbol{M}_{jj}(t)\mathbf{1}_{d_j} = \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} \boldsymbol{M}_{jk}(t)\mathbf{1}_{d_k} \tag{3.3.2}$$

contains the exit rate function out of macrostate j. This function is non-negative due to $\boldsymbol{M}_{jj}(t)$ being a sub-intensity matrix function. The last equality follows from the row sums of $\boldsymbol{M}(t)$ being zero.

In this paper, we give special attention to the case where $\boldsymbol{M}_{jk}$, $j, k \in \mathcal{J}$, $j \neq k$, is a matrix of rank one on the form

$$\boldsymbol{M}_{jk}(t) = \boldsymbol{\beta}_{jk}(t)\boldsymbol{\pi}_k(t), \tag{3.3.3}$$

where $\boldsymbol{\beta}_{jk}(t)$ is a d_j-dimensional non-negative column vector function and $\boldsymbol{\pi}_k(t)$ is a d_k-dimensional non-negative row vector function with $\boldsymbol{\pi}_k(t)\mathbf{1}_{d_k} = 1$. Here $\boldsymbol{\beta}_{jk}(t)$ provides the vector of jump rates from the microstates of macrostate j to the macrostate k, and $\boldsymbol{\pi}_k(t)$ denotes the initial distribution of $X_2(t)$ on $\{1, 2, \ldots, d_k\}$ just after a transition of $X_1(t)$ to k. In this case,

$$\boldsymbol{m}_j(t) = \sum_{\substack{k \in \mathcal{J} \\ k \neq j}} \boldsymbol{\beta}_{jk}(t). \tag{3.3.4}$$

Another way of writing (3.3.3) is

$$\mu_{(j,\tilde{j})(k,\tilde{k})}(t) = \beta_{(j,\tilde{j})k}(s)\pi_{(k,\tilde{k})}(s), \quad k \neq j.$$

We say that k has the reset property from state j. If all states k have the reset property from all states $j \neq k$, we simply say *the* reset property is satisfied.

Remark 3.3.1. Focusing only on the transition from macrostate j to macrostate k, we look at the following elements of $\boldsymbol{M}(t)$:

$$\begin{pmatrix} & \vdots & & \vdots & \\ \cdots & \boldsymbol{M}_{jj}(t) & \cdots & \boldsymbol{\beta}_{jk}(t)\boldsymbol{\pi}_k(t) & \cdots \\ & \vdots & \ddots & \vdots & \\ \cdots & \cdots & \cdots & \boldsymbol{M}_{kk}(t) & \cdots \\ & \vdots & & \vdots & \end{pmatrix}.$$

Here $\boldsymbol{\beta}_{jk}(t)$ is a column vector of exit rates from states $(j, \tilde{j})$ in $\{(j, 1), ..., (j, d_j)\}$. Therefore it seems natural to pair j and $\tilde{j}$, which explains the seemingly awkward indexation of elements of $\boldsymbol{\beta}_{jk}(t)$. Since a new microstate is picked independently of $(j, \tilde{j})$ from $\{(k, 1), ..., (k, d_k)\}$, records of where the process transitioned from are lost upon transition, which explains the term reset property. △

Let $\mathbb{F}^Z = \{\mathcal{F}^Z(t)\}_{t\geq 0}$ denote the natural filtration generated by the macrostate process Z. Since only Z is observed, the filtration $\mathbb{F}^Z$ represents the available information. We may, as previously, associate to Z a multivariate counting process N with components $N_{jk} = \{N_{jk}(t)\}_{t\geq 0}$ given by

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

as well as a marked point process $(T_n, Y_n)_{n=0}^{\infty}$ with T_n the n'th jump time of Z and $Y_n = Z(T_n)$; we use the convention $T_0 = 0$. Disregarding null-sets, the jump process Z, the multivariate counting process N, and the marked point process $(T_n, Y_n)_{n=0}^{\infty}$ generate the same information.

Although the microstate process $\boldsymbol{X}$ is Markovian, this is generally not the case for the macrostate process Z. In this paper, we derive distributional properties of Z by deriving distributional properties of the multivariate counting process N and the marked point process $(T_n, Y_n)_{n=0}^{\infty}$. We are especially interested in the special case where the reset property (3.3.3) holds across all states. Here it turns out that (Z, U) becomes Markovian, where U is the duration process defined in (3.2.2).

3.3.2 Payments

Having specified the probabilistic model, we now turn our attention to the insurance contract itself. Again, we denote by $B = \{B(t)\}_{t\geq 0}$ the payments, benefits less premiums, between the insured and the insurer. We suppose that B takes the form prescribed in (3.2.3). Furthermore, throughout the paper, we assume a maximal contract time $\eta > 0$ such that all sojourn payment rates and transition payments are zero after time η.

In the later stages of the paper, we add another layer of complications by turning to so-called scaled payments that appear in connection with policyholder behavior such as free-policy conversion and stochastic retirement. To be precise, here we furthermore consider payments $B^\rho = \{B^\rho(t)\}_{t\geq 0}$ given by

$$\mathrm{d}B^\rho(t) = \rho\big(\tau, Z(\tau-), Z(\tau)\big)^{1_{(\tau \leq t)}} \,\mathrm{d}B(t), \quad B^\rho(0) = B(0),$$

where τ is the exercise time of some policyholder option (modeled incidentally) and $0 < \rho(t,j,k) \leq 1$ is a suitable regular deterministic scaling factor.

The remainder of the paper now focuses on deriving distributional properties of the macrostate process Z, establishing computational schemes for relevant expected accumulated cash flows and prospective reserves, and finally relating these findings to existing models and methods in the life insurance literature.

3.4 Properties of Z

In this section, we derive some distributional properties of Z. In Subsection 3.4.1, we consider the general setup and derive the conditional finite-dimensional distributions of the marked point process $(T_n, Y_n)_{n\in\mathbb{N}_0}$ associated to Z as well as the predictable compensators of the multivariate counting processes N associated to Z. In Subsection 3.4.2, we impose the reset property, which we show, by applying the results of Subsection 3.4.1, leads to (Z,U) being Markovian.

3.4.1 General results

Since Z is generally not Markovian, we introduce

$$S_n = (T_0, Y_0, T_1, Y_1, \ldots, T_n, Y_n)$$

to keep track of the history of Z. Write

$$s_n = (0, 1, t_1, y_1, \ldots, t_n, y_n), \quad y_i \in \mathcal{J}, 0 < t_1 < t_2 < \ldots < t_n < \infty,$$

for a generic realization of S_n whenever $T_n < \infty$. Let

$$\bar{F}^{(n+1)}(t \,|\, s_n) = \mathbb{P}(T_{n+1} > t \,|\, S_n = s_n)$$

denote the conditional survival function of T_{n+1} given S_n, and let

$$G^{(n+1)}(k \,|\, s_n, t_{n+1}) = \mathbb{P}(Y_{n+1} = k \,|\, S_n = s_n, T_{n+1} = t_{n+1})$$

denote the conditional probability mass function of Y_{n+1} given (S_n, T_{n+1}). These quantities determine the distribution of Z. The following result provides a characterization of them within our setup.

Proposition 3.4.1. *The conditional finite-dimensional distributions of the marked point process $(T_n, Y_n)_{n=1}^{\infty}$ are given by*

$$\bar{F}^{(n+1)}(t|s_n) = \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \mathbf{1}_{d_{y_n}}, \quad t \geq t_n,$$

$$G^{(n+1)}(k|s_n, t_{n+1}) = \frac{\boldsymbol{\alpha}(s_n) \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_n k}(t_{n+1})\mathbf{1}_{d_k}}{\boldsymbol{\alpha}(s_n) \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{m}_{y_n}(t_{n+1})}, \quad k \neq y_n,$$

where the d_{y_n}-dimensional row vector $\boldsymbol{\alpha}(s_n)$ is given by

$$\boldsymbol{\alpha}(s_n) = \boldsymbol{\pi}_1(0) \prod_{\ell=0}^{n-1} \prod_{t_\ell}^{t_{\ell+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_\ell y_\ell}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_\ell y_{\ell+1}}(t_{\ell+1}).$$

Proof. Please refer to Appendix 3.A. □

Remark 3.4.2. The first statement of Proposition 3.4.1 corresponds to

$$T_{n+1} - T_n \,\big|\, S_n \sim \mathrm{IPH}\left(\frac{\boldsymbol{\alpha}(S_n)}{\boldsymbol{\alpha}(S_n)\mathbf{1}_{d_{Y_n}}}, \boldsymbol{M}_{Y_n Y_n}(T_n + \bullet)\right). \qquad \triangle$$

The compensators of the multivariate counting process associated to Z, which also determine the distribution of Z, are key quantities in the context of estimation and valuation. In our setup, they take the following form.

Theorem 3.4.3. *The counting process N_{jk} has $(\mathbb{F}^Z, \mathbb{P})$-compensator given by $\mathrm{d}\Lambda_{jk}(t) = \lambda_{jk}(t)\,\mathrm{d}t$ with*

$$\lambda_{jk}(t) = \sum_{n \in \mathbb{N}_0} 1_{(T_n,\, T_{n+1}]}(t) 1_{(Y_n = j)} \frac{\boldsymbol{\alpha}(S_{n-1}, T_n, j) \prod_{T_n}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\mathrm{d}x)}{\boldsymbol{\alpha}(S_{n-1}, T_n, j) \prod_{T_n}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\mathrm{d}x)\mathbf{1}_{d_j}} \boldsymbol{M}_{jk}(t)\mathbf{1}_{d_k}.$$

Proof. Please refer to Appendix 3.A. □

3.4.2 Reset property and semi-Markovianity

Suppose now that the reset property holds, that is $\boldsymbol{M}_{jk}(t)$ satisfies (3.3.3) for all $j \neq k$. We now make the following observations. From (3.3.3), we see that for $k \neq j$,

$$\boldsymbol{\pi}_j(t) \prod_t^s (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)\,\boldsymbol{\beta}_{jk}(s)$$

is a 1×1-dimensional matrix, implying it cancels if appearing in both the numerator and denominator of a fraction. In particular,

$$\frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} = \frac{\boldsymbol{\pi}_{y_n}(t_n)}{\boldsymbol{\pi}_{y_n}(t_n)\mathbf{1}_{d_{y_n}}} = \boldsymbol{\pi}_{y_n}(t_n). \tag{3.4.1}$$

Combined with the results of Subsection 3.4.1, this yields the following corollaries.

Corollary 3.4.4. *Assume that* (3.3.3) *holds. Then the conditional finite-dimensional distributions of the marked point process* $(T_n, Y_n)_{n=1}^{\infty}$ *are given by*

$$\bar{F}^{(n+1)}(t|s_n) = \boldsymbol{\pi}_{y_n}(t_n) \prod_{t_n}^{t} (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x)\,\mathbf{1}_{d_{y_n}}, \quad t \geq t_n,$$

$$G^{(n+1)}(k|s_n, t_{n+1}) = \frac{\boldsymbol{\pi}_{y_n}(t_n) \prod_{t_n}^{t_{n+1}} (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x)\,\boldsymbol{\beta}_{y_n y_{n+1}}(t_{n+1})}{\boldsymbol{\pi}_{y_n}(t_n) \prod_{t_n}^{t_{n+1}} (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x)\,\boldsymbol{m}_{y_n}(t_{n+1})}, \quad k \neq y_n,$$

where $\boldsymbol{m}_{y_n}(t_{n+1})$ *is obtained from* (3.3.4)*.*

Remark 3.4.5. The first statement of Corollary 3.4.4 corresponds to

$$T_{n+1} - T_n \,\big|\, S_n \sim \mathrm{IPH}\big(\boldsymbol{\pi}_{Y_n}(T_n),\, \boldsymbol{M}_{Y_n Y_n}(T_n + \bullet)\big). \qquad \triangle$$

Corollary 3.4.6. *Assume that* (3.3.3) *holds. Then the counting process* N_{jk} *has* $(\mathbb{F}^Z, \mathbb{P})$*-compensator given by* $\mathrm{d}\Lambda_{jk}(t) = \lambda_{jk}(t)\,\mathrm{d}t$ *with*

$$\lambda_{jk}(t) = \sum_{n \in \mathbb{N}_0} 1_{(T_n,\, T_{n+1}]}(t) 1_{(Y_n = j)} \frac{\boldsymbol{\pi}_j(T_n) \prod_{T_n}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)}{\boldsymbol{\pi}_j(T_n) \prod_{T_n}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)\mathbf{1}_{d_j}} \boldsymbol{\beta}_{jk}(t).$$

The results show that the general path dependence of Z through $\boldsymbol{\alpha}(S_n)$ is significantly reduced whenever (3.3.3) is imposed. We may actually write

$$\lambda_{jk}(t) = 1_{(Z(t-)=j)} \frac{\boldsymbol{\pi}_j(t - U(t-)) \prod_{t-U(t-)}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)}{\boldsymbol{\pi}_j(t - U(t-)) \prod_{t-U(t-)}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)\mathbf{1}_{d_j}} \boldsymbol{\beta}_{jk}(t),$$

which shows that the macrostate process Z is a time-inhomogeneous semi-Markov process with transition rates $\nu_{jk}(t,u)$, $j \neq k$, which are functions of both time and duration, given by

$$\nu_{jk}(t,u) = \frac{\boldsymbol{\pi}_j(t-u) \prod_{t-u}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)}{\boldsymbol{\pi}_j(t-u) \prod_{t-u}^{t} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\,\mathrm{d}x)\mathbf{1}_{d_j}} \boldsymbol{\beta}_{jk}(t).$$

Conversely, the class of aggregate Markov models is quite flexible. In light also of the many prized denseness results for phase type distributions, we therefore conjecture that the class of aggregate Markov models with the reset property is dense in the class of all (smooth) semi-Markovian models. That is, any (smooth) semi-Markovian model can be approximated arbitrarily well by an aggregate Markov model with the reset property simply by letting the number of microstates increase to infinity. The clarification of this conjecture is outside the scope of this paper, but it nevertheless constitutes an interesting research direction.

3.5 Valuation

In this section, we consider the valuation of the life insurance liabilities corresponding to the payment process B. In Subsection 3.5.1, we provide expressions for the expected accumulated cash flows and, hereby, the prospective reserves. In particular, we provide matrix representations that are useful in implementing the models. The expected accumulated cash flows are composed of conditional occupation probabilities, for which we derive formulas in Subsection 3.5.2. Special emphasis is given to the semi-Markovian case of Subsection 3.4.2. Finally, in Subsection 3.5.3 and Subsection 3.5.4, we investigate the impact of duration-independent payments and the inclusion of policyholder options, respectively.

Throughout the section, the time value of money is described by a deterministic and suitably regular interest rate $r(t)$. As long as financial and insurance risks are assumed independent, the extension to stochastic interest rates is straightforward.

3.5.1 General results

The expected accumulated cash flow $A(t,s)$ valued at time t is given by,

$$A(t,s) = \mathbb{E}\big[B(s) - B(t) \,\big|\, \mathcal{F}^Z(t)\big], \quad s \geq t,$$

confer with Definition 2.2 in Buchardt, Furrer, and Steffensen (2019), so that the prospective reserve reads

$$V(t) = \mathbb{E}\bigg[\int_t^\eta e^{-\int_t^s r(v)\,\mathrm{d}v}\,\mathrm{d}B(s)\,\bigg|\,\mathcal{F}^Z(t)\bigg] = \int_t^\eta e^{-\int_t^s r(v)\,\mathrm{d}v} A(t,\mathrm{d}s). \tag{3.5.1}$$

If the payments are on the form (3.2.3), then

$$A(t,\mathrm{d}s) = \sum_{\boldsymbol{j}\in E}\int_0^{U(t)+s-t} p_{\boldsymbol{j}}(t,s,\mathrm{d}z)\bigg(b_j(s,z) + \sum_{\substack{\boldsymbol{k}\in E\\ k\neq j}} b_{jk}(s,z)\mu_{\boldsymbol{jk}}(s)\bigg)\mathrm{d}s,$$

where the conditional occupation probabilities are given by

$$p_{\boldsymbol{j}}(t,s,z) = \mathbb{P}\big(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) \leq z \,\big|\, \mathcal{F}^Z(t)\big). \tag{3.5.2}$$

If the reset property (3.3.3) is also satisfied, then Z is a time-inhomogeneous semi-Markovian process and thus $A(t,s) = \sum_{i\in\mathcal{J}} 1_{(Z(t)=i)} A_{i,U(t)}(t,s)$, where

$$A_{i,u}(t,s) = \mathbb{E}\big[B(s) - B(t) \,\big|\, Z(t) = i, U(t) = u\big].$$

Furthermore, it holds that

$$A_{i,u}(t,\mathrm{d}s) = \sum_{\boldsymbol{j}\in E}\int_0^{u+s-t} p_{i\boldsymbol{j}}(t,u,s,\mathrm{d}z)\bigg(b_j(s,z) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b_{jk}(s,z)\beta_{\boldsymbol{j}k}(s)\bigg)\mathrm{d}s,$$

where the transition probabilities are given by

$$p_{i\boldsymbol{j}}(t,u,s,z) = \mathbb{P}\big(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) \leq z \,\big|\, Z(t) = i, U(t) = u). \tag{3.5.3}$$

For implementation purposes, it may be beneficial to use matrix representations of the expected accumulated cash flow $A(t,s)$ following along the lines of Bladt, Asmussen, and Steffensen (2020), since it allows for more compact and direct computations.

In the general case where the aforementioned reset property is not satisfied, the process Z is non-Markovian, so it is not sensible to form a transition probability matrix function in the usual way. Instead, we form a $\bar{d}$-dimensional vector function according to

$$\boldsymbol{p}(t,s,\mathrm{d}z) = \big\{p_{\boldsymbol{j}}(t,s,\mathrm{d}z)\big\}_{\boldsymbol{j}\in E}. \tag{3.5.4}$$

In regards to payments and transition rates, however, the fact that $\boldsymbol{X}$ is assumed to be Markovian allows us to follow more closely the approach of Bladt, Asmussen, and Steffensen (2020). In the present setup, we have a set of sojourn payment rates and transition payments that are all identical across microstates (of the same macrostate). Hence, the $\bar{d}$-dimensional vector of sojourn payment rates on the micro level is given by

$$\boldsymbol{b}(t,u) = (\boldsymbol{b}_1(t,u), ..., \boldsymbol{b}_J(t,u)), \tag{3.5.5}$$

where $\boldsymbol{b}_j(t,u) = b_j(t,u)\mathbf{1}_{d_j}$. The matrices of transition payments must, in a similar fashion, be identical across microstates (of the same macrostate), so that the transition payment matrix function on the micro level is given by

$$\boldsymbol{B}(t,u) = \begin{pmatrix} \boldsymbol{B}_{11}(t,u) & \boldsymbol{B}_{12}(t,u) & \cdots & \boldsymbol{B}_{1J}(t,u) \\ \boldsymbol{B}_{21}(t,u) & \boldsymbol{B}_{22}(t,u) & \cdots & \boldsymbol{B}_{2J}(t,u) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B}_{J1}(t,u) & \boldsymbol{B}_{J2}(t,u) & \cdots & \boldsymbol{B}_{JJ}(t,u) \end{pmatrix}, \tag{3.5.6}$$

where $\boldsymbol{B}_{ij}(t,u)$, $i,j \in \mathcal{J}$, $j \neq i$, is a $d_i \times d_j$-dimensional matrix with $b_{ij}(t,u)$ in all entries, and $\boldsymbol{B}_{ii}(t,u) = \mathbf{0}$ is a $d_i \times d_i$-dimensional matrix of zeroes. Based hereon, we define the reward matrix function as

$$\boldsymbol{R}(t,u) = \boldsymbol{\Delta}(\boldsymbol{b}(t,u)) + \boldsymbol{M}(t) \bullet \boldsymbol{B}(t,u), \tag{3.5.7}$$

where $\bullet$ denotes the Schur product, that is $(\boldsymbol{A} \bullet \boldsymbol{B})_{ij} = A_{ij}B_{ij}$, and $\boldsymbol{\Delta}(\boldsymbol{b})$ is a diagonal matrix with the vector $\boldsymbol{b}$ as diagonal. This is similar to equations (3.8)–(3.11) in Bladt, Asmussen, and Steffensen (2020).

The expected accumulated cash flow $A(t,s)$ may then be seen to have the following matrix representation:

$$A(t, \mathrm{d}s) = \int_0^{U(t)+s-t} \boldsymbol{p}(t,s,\mathrm{d}z)\boldsymbol{R}(s,z)\mathbf{1}_{\bar{d}}\,\mathrm{d}s, \tag{3.5.8}$$

where the original sums over the state space E are reduced to matrix multiplications.

In the case of the reset property (3.3.3), the semi-Markovianity of Z implies that it suffices to consider the transition probabilities of (3.5.3). Thus, it is sensible to form a $J \times \bar{d}$-dimensional matrix function according to

$$\boldsymbol{p}(t,u,s,\mathrm{d}z) = \{p_{i\boldsymbol{j}}(t,u,s,\mathrm{d}z)\}_{i\in\mathcal{J},\boldsymbol{j}\in E}.$$

Similarly, we may form the J-dimensional vector

$$\boldsymbol{A}_u(t,s) = (A_{1,u}(t,s), \ldots, A_{J,u}(t,s))$$

of state-wise expected accumulated cash flows, which then can be calculated as follows:

$$\boldsymbol{A}_u(t, \mathrm{d}s) = \int_0^{u+s-t} \boldsymbol{p}(t, u, s, \mathrm{d}z)\boldsymbol{R}(s, z)\boldsymbol{1}_{\bar{d}}\,\mathrm{d}s, \tag{3.5.9}$$

where the reward matrices $\boldsymbol{R}(t, u)$ of (3.5.7) are modified according to (3.3.3)–(3.3.4). We can also cast (3.5.9) as $\boldsymbol{A}_u(t, \mathrm{d}s) = \boldsymbol{a}_u(t, s)\,\mathrm{d}s$, where then

$$\boldsymbol{a}_u(t, s) = \int_0^{u+s-t} \boldsymbol{p}(t, u, s, \mathrm{d}z)\boldsymbol{R}(s, z)\boldsymbol{1}_{\bar{d}}$$

is a vector of state-wise expected cash flows.

3.5.2 Conditional occupation and transition probabilities

We now provide calculation schemes for the conditional occupation and transition probabilities. Rather than working directly with these quantities, it turns out to be fruitful to focus instead on

$$\begin{aligned}
\bar{p}_{\boldsymbol{\jmath}}(t, s, z) &= \mathbb{P}\big(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,\big|\, \mathcal{F}^Z(t)\big), \\
\bar{p}_{i\boldsymbol{\jmath}}(t, u, s, z) &= \mathbb{P}\big(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,\big|\, Z(t) = i, U(t) = u\big),
\end{aligned}$$

which suffices since

$$\begin{aligned}
\bar{p}_{\boldsymbol{\jmath}}(t, s, \mathrm{d}z) &= -p_{\boldsymbol{\jmath}}(t, s, \mathrm{d}z), \\
\bar{p}_{i\boldsymbol{\jmath}}(t, u, s, \mathrm{d}z) &= -p_{i\boldsymbol{\jmath}}(t, u, s, \mathrm{d}z).
\end{aligned}$$

In the following, we require the $\bar{d} \times d_j$-dimensional matrices

$$\boldsymbol{E}_j = \sum_{\widetilde{j}=1}^{d_j} \boldsymbol{e}_{\boldsymbol{\jmath}} \boldsymbol{e}_{\widetilde{j}}',$$

where $\boldsymbol{e}_{\widetilde{j}}$ is a d_j-dimensional column with a one in entry $\widetilde{j}$ and otherwise zeroes, and $\boldsymbol{e}_{\boldsymbol{\jmath}}$ is a $\bar{d}$-dimensional column vector with a one in entry $d_1 + \cdots + d_{j-1} + \widetilde{j}$ and otherwise zeroes. Here and in the following, primes denotes matrix transposition. The entries of $\boldsymbol{E}_j$ are zero, except in the j'th block row, where they consist of the d_j-dimensional identity matrix. Roughly speaking, they allow us to extend a distribution on microstates in a single macrostate to the whole state space E (and vice versa).

Theorem 3.5.1. *It holds that*

$$\bar{p}_{\boldsymbol{\jmath}}(t, s, z) = \sum_{n \in \mathbb{N}_0} 1_{[T_n, T_{n+1})}(t)\widetilde{p}_{\boldsymbol{\jmath}}(t, s, z; S_n),$$

where the auxiliary quantities $\widetilde{p}_{\boldsymbol{j}}(t,s,z;s_n)$ *are zero for* $t_n \geq s-z$ *and*

$$\widetilde{p}_{\boldsymbol{j}}(t,s,z;s_n) = \frac{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{t}(\boldsymbol{I}+\boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\boldsymbol{E}'_{y_n}}{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{t}(\boldsymbol{I}+\boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\boldsymbol{1}_{d_{y_n}}} \prod_{t}^{(s-z)\vee t}(\boldsymbol{I}+\boldsymbol{M}(x)\mathrm{d}x)\boldsymbol{E}_j \prod_{(s-z)\vee t}^{s}(\boldsymbol{I}+\boldsymbol{M}_{jj}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}}$$

for $t_n < s-z$.

Note that

$$1_{(t_n < s-z \leq t)}\widetilde{p}_{\boldsymbol{j}}(t,s,z;s_n) = 1_{(y_n=j)}\frac{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{s}(\boldsymbol{I}+\boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}}}{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{t}(\boldsymbol{I}+\boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{1}_{d_{y_n}}}. \tag{3.5.10}$$

Proof. Please refer to Appendix 3.A. □

If the reset property (3.3.3) is satisfied, in which case Z is a time-inhomogeneous semi-Markovian process, we can use (3.4.1) to immediately obtain the following corollary.

Corollary 3.5.2. *Assume* (3.3.3) *holds. Then* $\bar{p}_{i\boldsymbol{j}}(t,u,s,z)$ *is zero for* $t-u \geq s-z$ *and*

$$\bar{p}_{i\boldsymbol{j}}(t,u,s,z) = \frac{\boldsymbol{\pi}_i(t-u)\prod_{t-u}^{t}(\boldsymbol{I}+\boldsymbol{M}_{ii}(x)\mathrm{d}x)\boldsymbol{E}'_i}{\boldsymbol{\pi}_i(t-u)\prod_{t-u}^{t}(\boldsymbol{I}+\boldsymbol{M}_{ii}(x)\mathrm{d}x)\boldsymbol{1}_{d_i}} \prod_{t}^{(s-z)\vee t}(\boldsymbol{I}+\boldsymbol{M}(x)\mathrm{d}x)\boldsymbol{E}_j \prod_{(s-z)\vee t}^{s}(\boldsymbol{I}+\boldsymbol{M}_{jj}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}}$$

for $t-u < s-z$.

We note that $z \mapsto \bar{p}_{i\boldsymbol{j}}(t,u,s,z)$ is continuous on $[0,u+s-t)$ and actually constant on $[s-t,u+s-t)$. If $i \neq j$, then the continuity extends to $[0,u+s-t]$, while

$$\Delta\bar{p}_{i\boldsymbol{i}}(t,u,s,u+s-t) = -\lim_{h\downarrow 0}\bar{p}_{i\boldsymbol{i}}(t,u,s,s-t-h).$$

The fact that $z \mapsto \bar{p}_{i\boldsymbol{j}}(t,u,s,z)$ is constant on $[s-t,u+s-t)$ may be utilized to reduce the computational load when calculating the expected accumulated cash flows.

Algorithm 3.1 Computation of expected cash flows in an aggregate Markov model with the reset property.

Input: Current time $t \in [0, \eta)$, current duration $u \in [0, t]$, and a grid $\mathcal{T} : t = t_0 < t_1 < \cdots < t_n = \eta$ on the interval $[t, \eta]$.

1) Calculate initial conditional distributions at time t:

$$\boldsymbol{\gamma}_i(t, u) = \frac{\boldsymbol{\pi}_i(t-u) \prod_{t-u}^{t} (\boldsymbol{I} + \boldsymbol{M}_{ii}(x)\mathrm{d}x)\, \boldsymbol{E}_i'}{\boldsymbol{\pi}_i(t-u) \prod_{t-u}^{t} (\boldsymbol{I} + \boldsymbol{M}_{ii}(x)\mathrm{d}x)\, \boldsymbol{1}_{d_i}}, \qquad i \in \mathcal{J},$$

$$\boldsymbol{\gamma}(t, u) = (\boldsymbol{\gamma}_1(t, u), \ldots, \boldsymbol{\gamma}_J(t, u))'.$$

2) Compute transition probabilities for the Markovian state process $\boldsymbol{X}$,

$$\boldsymbol{P}(t, t_\ell) = \prod_{t}^{t_\ell} (\boldsymbol{I} + \boldsymbol{M}(x)\mathrm{d}x), \qquad \ell \in \{1, \ldots, n\},$$

by solving Kolmogorov's forward differential equation on $\mathcal{T}$.

3) For $\ell \in \{1, \ldots, n\}$:

i) Compute state-wise stay probabilities until time t_ℓ:

$$\bar{\boldsymbol{P}}_{jj}(t_{\ell'}, t_\ell) = \prod_{t_{\ell'}}^{t_\ell} (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\mathrm{d}x), \qquad t_{\ell'} \in \mathcal{T},\ \ell' \leq \ell,\ j \in \mathcal{J},$$

$$\bar{\boldsymbol{P}}(t_{\ell'}, t_\ell) = \boldsymbol{\Delta}\big((\bar{\boldsymbol{P}}_{11}(t_{\ell'}, t_\ell), \ldots, \bar{\boldsymbol{P}}_{JJ}(t_{\ell'}, t_\ell)\big),$$

by solving Kolmogorov's backward differential equation on $\mathcal{T}$ starting at t_ℓ.

ii) Calculate the vector of state-wise expected cash flows for time t_ℓ:

$$\boldsymbol{a}_u(t, t_\ell) = \boldsymbol{\gamma}(t, u)\bigg(\int_t^{t_\ell} \boldsymbol{P}(t, v)\widetilde{\boldsymbol{M}}(v)\bar{\boldsymbol{P}}(v, t_\ell)\boldsymbol{R}(t_\ell, t_\ell - v)\boldsymbol{1}_{\bar{d}}\, \mathrm{d}v \\ - \bar{\boldsymbol{P}}(t, t_\ell)\boldsymbol{R}(t_\ell, u + t_\ell - t)\boldsymbol{1}_{\bar{d}}\bigg),$$

using numerical integration methods on the grid $\mathcal{T}$ for the integral, and where

$$\widetilde{\boldsymbol{M}}_j(v) = \boldsymbol{M}(v)\boldsymbol{E}_j - \boldsymbol{E}_j\boldsymbol{M}_{jj}(v), \qquad j \in \mathcal{J},$$

$$\widetilde{\boldsymbol{M}}(v) = \Big(\widetilde{\boldsymbol{M}}_1(v), \ldots, \widetilde{\boldsymbol{M}}_J(v)\Big).$$

Output*: For each $\ell \in \{1, \ldots, n\}$, a vector of state-wise expected cash flows $\boldsymbol{a}_u(t, t_\ell)$.*

We conclude this subsection by presenting the above algorithm for the computation of expected cash flows in models with the reset property. The computational scheme is similar to the algorithm for general semi-Markov models based on Kolmogorov's forward integro-differential equation proposed in Buchardt, Møller, and Schmidt (2015, Section 3). Both algorithms employ a two-dimensional time and duration grid, and one would therefore expect the computational loads to be comparable.

3.5.3 Duration-independent payments

We now consider the simplifications arising from duration-independent payments, that is, when

$$b_j(t,u) = b_j(t) \quad \text{and} \quad b_{jk}(t,u) = b_{jk}(t), \tag{3.5.11}$$

or, equivalently,

$$\boldsymbol{b}(t,u) = \boldsymbol{b}(t) \quad \text{and} \quad \boldsymbol{B}(t,u) = \boldsymbol{B}(t).$$

In this case,

$$\begin{aligned} A(t,\mathrm{d}s) &= \sum_{\boldsymbol{j}\in E} \bar{p}_{\boldsymbol{j}}(t,s,0)\Bigg(b_j(s) + \sum_{\substack{\boldsymbol{k}\in E\\ k\neq j}} b_{jk}(s)\mu_{\boldsymbol{jk}}(s)\Bigg)\mathrm{d}s \\ &= \bar{\boldsymbol{p}}(t,s,0)\boldsymbol{R}(s)\boldsymbol{1}_{\bar{d}}\,\mathrm{d}s, \end{aligned}$$

where

$$\begin{aligned} \bar{\boldsymbol{p}}(t,s,0) &= \left\{\bar{p}_{\boldsymbol{j}}(t,s,0)\right\}_{\boldsymbol{j}\in E}, \\ \boldsymbol{R}(t) &= \boldsymbol{\Delta}(\boldsymbol{b}(t)) + \boldsymbol{M}(t) \bullet \boldsymbol{B}(t). \end{aligned}$$

According to Theorem 3.5.1,

$$\bar{p}_{\boldsymbol{j}}(t,s,0) = \sum_{n\in\mathbb{N}_0} 1_{[T_n,\,T_{n+1})}(t)\widetilde{p}_{\boldsymbol{j}}(t,s,0;S_n),$$

$$\widetilde{p}_{\boldsymbol{j}}(t,s,0;s_n) = \frac{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{t}(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{E}'_{y_n}}{\boldsymbol{\alpha}(s_n)\prod_{t_n}^{t}(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{1}_{d_{y_n}}}\prod_{t}^{s}(\boldsymbol{I} + \boldsymbol{M}(x)\mathrm{d}x)\boldsymbol{e}_{\boldsymbol{j}}.$$

If the reset property (3.3.3) is satisfied, then we are rather interested in $\boldsymbol{A}_u(t,s)$, which subject to (3.5.11) reads

$$\boldsymbol{A}_u(t,\mathrm{d}s) = \bar{\boldsymbol{p}}(t,u,s,0)\boldsymbol{R}(s)\boldsymbol{1}_{\bar{d}}\,\mathrm{d}s,$$

where

$$\bar{\boldsymbol{p}}(t,u,s,0) = \left\{\bar{p}_{i\boldsymbol{j}}(t,u,s,0)\right\}_{i\in\mathcal{J},\boldsymbol{j}\in E}.$$

Furthermore,

$$\bar{p}_{i\boldsymbol{j}}(t,u,s,0) = \frac{\boldsymbol{\pi}_i(t-u)\prod_{t-u}^{t}\left(\boldsymbol{I}+\boldsymbol{M}_{ii}(x)\mathrm{d}x\right)\boldsymbol{E}_i'}{\boldsymbol{\pi}_i(t-u)\prod_{t-u}^{t}\left(\boldsymbol{I}+\boldsymbol{M}_{ii}(x)\mathrm{d}x\right)\mathbf{1}_{d_i}}\prod_{t}^{s}(\boldsymbol{I}+\boldsymbol{M}(x)\mathrm{d}x)\boldsymbol{e}_{\boldsymbol{j}}.$$

This should lead to a significant reduction in computational load since the above simplification allows one to adapt Algorithm 3.1 to employ only a one-dimensional time grid. For general semi-Markov models, where the computation of transition probabilities relies on Kolmogorov's forward integro-differential equation, such a simplification is not possible. It should be noted, however, that the computation of the term

$$\prod_{t}^{s}(\boldsymbol{I}+\boldsymbol{M}(x)\mathrm{d}x)$$

might still be rather burdensome if $\bar{d}$ is large. To conclude, if the number of microstates per macrostates is not too large, aggregate Markov models might hold a competitive advantage over general semi-Markov models if the payments of interest are duration-independent.

The above discussion relates to duration-independent payments. However, it is also applicable to certain crude duration-dependent payments. This is partly illustrated by the numerical example in Section 3.6, where we consider a contract stipulating a waiting period.

3.5.4 Policyholder behavior

We now extend the results of Subsections 3.5.1–3.5.2 to include incidental policyholder behavior such as free-policy conversion and expedited or postponed retirement. The inclusion of policyholder options is quite popular in the life insurance literature, confer with Henriksen et al. (2014), Buchardt and Møller (2015), Buchardt, Møller, and Schmidt (2015), and Gad and Nielsen (2016), especially for Markov chains. General insights based on change of measure techniques were recently provided in Furrer (2022). In the following, we adapt the general methods and results of Furrer (2022) to our setting.

Suppose that the macrostates $\mathcal{J}$ can be decomposed as

$$\mathcal{J} = \mathcal{J}_0 \cup \mathcal{J}_1 \cup \{\nabla\},$$

with $1 \in \mathcal{J}_0$ and where the transition intensity matrix function $\boldsymbol{M}(t)$ of the microstate process $\boldsymbol{X}$ is composed of block matrix functions satisfying

$$\begin{aligned} \boldsymbol{M}_{jk}(t) &= \boldsymbol{0}, && j \in \mathcal{J}_1, k \in \mathcal{J}_0, \\ \boldsymbol{M}_{j\nabla}(t) &= \boldsymbol{0}, && j \in \mathcal{J}. \end{aligned}$$

In that case, the macrostate process Z almost surely never hits ∇, and, upon entering the states $\mathcal{J}_1$, the process never returns to $\mathcal{J}_0$. Recall that $Z(0) \equiv 1$, so the process starts in $\mathcal{J}_0$. We may thus interpret $\mathcal{J}_0$ as the states of the insured prior to exercising their policyholder option and $\mathcal{J}_1$ as the states of the insured after exercising the option. (The role of the 'dummy' state ∇ will be clear later.) The first hitting time of $\mathcal{J}_1$, given by

$$\tau = \inf\{t > 0 : Z(t) \in \mathcal{J}_1\},$$

then constitutes the exercise time of the option. Since Z almost surely never hits ∇, we may as well take $b_\nabla(s,u) = 0$, $b_{j\nabla}(s,u) = 0$, and $b_{\nabla k}(s,u) = 0$. At exercise, the original contractual payments are scaled with the factor $\rho(\tau, Z(\tau-), Z(\tau))$, where $0 < \rho(t,j,k) \leq 1$ is some suitably regular deterministic function. The payment process of interest $B^\rho = \{B^\rho(t)\}_{t\geq 0}$ thus takes the form

$$\mathrm{d}B^\rho(t) = \rho\big(\tau, Z(\tau-), Z(\tau)\big)^{1_{(\tau\leq t)}} \, \mathrm{d}B(t), \quad B^\rho(0) = B(0).$$

The scaling factor is typically selected as to maintain actuarial equivalence with respect to a safe-side valuation basis, the so-called technical basis; we just consider it given. The corresponding expected accumulated cash flow $A^\rho(t,s)$ valued at time t is

$$A^\rho(t,s) = \mathbb{E}[B^\rho(s) - B^\rho(t) | \mathcal{F}^Z(t)] = \mathbb{E}\left[\int_t^s \rho\big(\tau, Z(\tau-), Z(\tau)\big)^{1_{(\tau\leq u)}} \mathrm{d}B(u) \middle| \mathcal{F}^Z(t) \right].$$

The following result is a consequence of Furrer (2022, Theorem 3.6 and Proposition 3.10).

Proposition 3.5.3. *It holds that*

$$A^\rho(t,s) = \widehat{\mathbb{E}}[B(s) - B(t) \,|\, \mathcal{F}^Z(t)] \rho\big(\tau, Z(\tau-), Z(\tau)\big)^{1_{(\tau\leq t)}},$$

where $\widehat{\mathbb{E}}$ denotes expectation with respect to another probability measure $\widehat{\mathbb{P}}$. Furthermore, the $(\mathbb{F}^Z, \widehat{\mathbb{P}})$-compensators of the counting processes are given by

$$\begin{aligned} \mathrm{d}\widehat{\Lambda}_{jk}(t) &= \rho(t,j,k)\, \mathrm{d}\Lambda_{jk}(t), && j \in \mathcal{J}_0, k \in \mathcal{J}_1, \\ \mathrm{d}\widehat{\Lambda}_{j\nabla}(t) &= \sum_{k\in\mathcal{J}_1} (1 - \rho(t,j,k))\, \mathrm{d}\Lambda_{jk}(t), && j \in \mathcal{J}_0, \\ \mathrm{d}\widehat{\Lambda}_{\nabla k}(t) &= 0, && k \in \mathcal{J}, k \neq \nabla, \\ \mathrm{d}\widehat{\Lambda}_{jk}(t) &= \mathrm{d}\Lambda_{jk}(t), && \textit{otherwise}. \end{aligned}$$

Recall that the compensators determine the distribution of Z. Thus from the expressions for the compensators obtained in Theorem 3.4.3 and the above proposition, we find that Z under $\widehat{\mathbb{P}}$ follows an aggregate Markov model with transition intensity matrix function $\widehat{\boldsymbol{M}}(t)$ composed of block matrix functions

$$
\begin{aligned}
\widehat{\boldsymbol{M}}_{jk}(t) &= \rho(t,j,k)\boldsymbol{M}_{jk}(t), && j \in \mathcal{J}_0, k \in \mathcal{J}_1,\\
\widehat{\boldsymbol{M}}_{j\nabla}(t) &= \sum_{k\in\mathcal{J}_1}(1-\rho(t,j,k))\boldsymbol{M}_{jk}(t)\mathbf{1}_{d_k}, && j \in \mathcal{J}_0,\\
\widehat{\boldsymbol{M}}_{\nabla k}(t) &= \mathbf{0}, && k \in \mathcal{J}, k \neq \nabla,\\
\widehat{\boldsymbol{M}}_{jk}(t) &= \boldsymbol{M}_{jk}(t), && \text{otherwise.}
\end{aligned}
$$

Furthermore, if the reset property (3.3.3) is satisfied under $\mathbb{P}$, this is also the case under $\widehat{\mathbb{P}}$. All in all, according to Proposition 3.5.3 the expected accumulated cash flow

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

and thus also the expected accumulated cash flow $A^\rho(t,s)$, can be calculated using the formulas of Subsections 3.5.1–3.5.2, but with $\boldsymbol{M}(t)$ replaced by $\widehat{\boldsymbol{M}}(t)$.

3.6 Numerical example

We conclude the paper by presenting a numerical example that serves to illustrate the methods presented in Section 3.5. The probabilistic models, described by transition rates on the micro level, are taken from the numerical example in Ahmad and Bladt (2022a), where aggregate Markov models corresponding to Figure 3.1 with the reset property are fitted to simulated data on a macro level for different numbers of disability microstates, d_2. The simulations are based on a (smooth) semi-Markovian disability model employed by a large Danish life insurance company that has been reported to and published by the Danish Financial Supervisory Authority. The only duration effects present in this model concern the rates from the disability state, which also explains why we do not add extra microstates to the active macrostate. The rates from the disability state are, at least after some months, decreasing as functions of duration. We refer to the numerical example in Ahmad and Bladt (2022a) for further details.

In the following, the analysis of Ahmad and Bladt (2022a) is extended with calculations and comparisons of state-wise expected cash flows and prospective reserves. We focus on a coverage which admits duration-dependent payments, namely a disability coverage with a waiting period. To be specific, we consider a male of age $t = 40$ years with a disability annuity of rate 1 per year, starting 3 months after the onset of disability, but only until retirement at age 65. The only

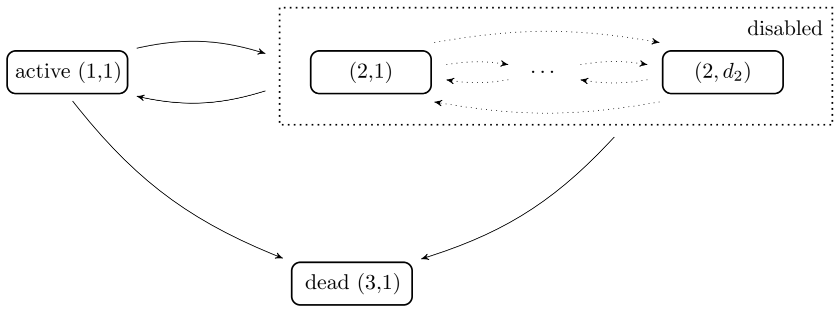

Figure 3.1: *Disability model with d_2 unobservable disability microstates.*

non-zero payments function is thus $b_2(s, z)$, which reads

$$b_2(s, z) = 1_{(s<65)}1_{(z>1/4)},$$

and we may therefore set $\eta = 65$.

We emphasize that this particular simple type of duration dependence allows for simplifications in the computation schemes similar to those from the duration-independent case of Subsection 3.5.3. Indeed, the vector of state-wise expected cash flows now reads

$$\boldsymbol{a}_u(40, s) = 1_{(u+s-40>1/4)}\bar{\boldsymbol{p}}(40, s, u, 1/4)\boldsymbol{E}_2\mathbf{1}_{d_2},$$

where the elements of $\bar{\boldsymbol{p}}(40, s, u, 1/4)$ can be calculated using Corollary 3.5.2, confer also with Algorithm 3.1. Since we only need the transition probabilities at a single (and rather small) end duration $z = 1/4$, the computational complexity is comparable with that of the duration-independent case, where only $z = 0$ is needed.

The corresponding vector of state-wise prospective reserves is obtained by discounting and accumulating the vector of state-wise expected cash flows:

$$\boldsymbol{V}_u(40) = \int_{40}^{65} e^{-\int_{40}^{s} r(v)\,\mathrm{d}v}\boldsymbol{a}_u(40, s)\,\mathrm{d}s.$$

For the interest rate, we use the forward rate curve published on the 3rd of November, 2022, by the Danish Financial Supervisory Authority. We are implicitly assuming that the financial market is independent of the state process of the insured.

We calculate the vectors of state-wise expected cash flows and corresponding prospective reserves across initial states and durations and for the various fits. We also calculate these quantities for the underlying true semi-Markov model, where we need to use Kolmogorov's forward integro-differential equation of Buchardt, Møller, and Schmidt (2015, Section 3), since this model is not an aggregate Markov model. Figure 3.2 shows the resulting expected cash flows in the disability state for initial durations $u = 0$ and $u = 1$, while Figure 3.3 shows the prospective disability reserves as functions of the duration since the onset of disability.

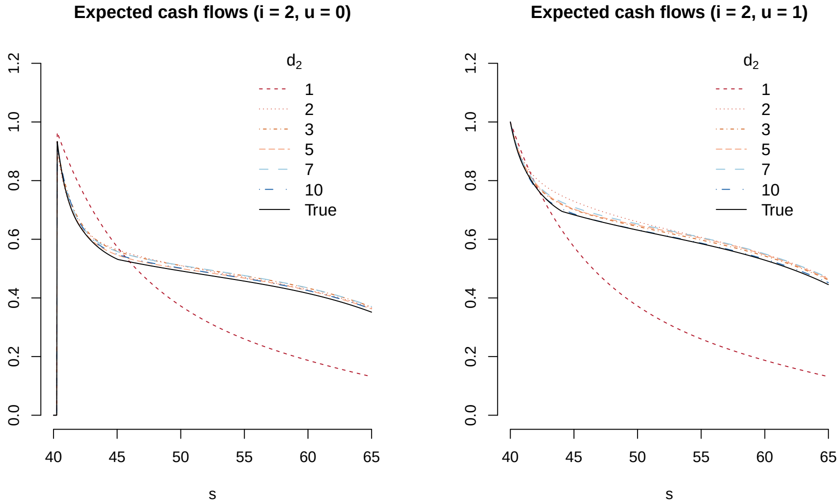

Figure 3.2: *Expected cash flow $s \mapsto a_{2,u}(40, s)$ in the disability state with initial durations $u = 0$ (left) and $u = 1$ (right) for different numbers of disability microstates, d_2, along with the true expected accumulated cash flows. The case $d_2 = 1$ corresponds to a Markov chain.*

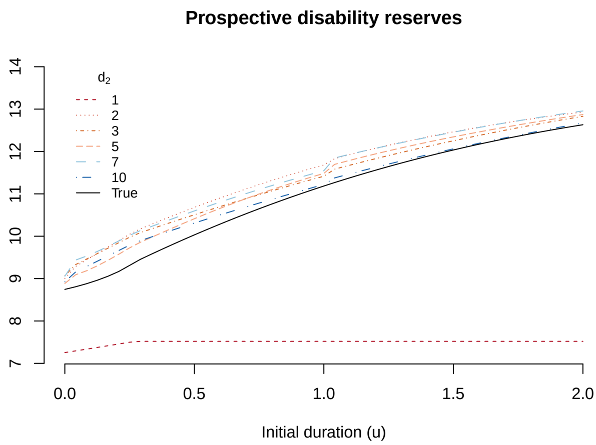

Figure 3.3: *Prospective disability reserve as a function of duration since onset of disability, $u \mapsto V_{2,u}(40)$, for different numbers of disability microstates, d_2, along with true prospective reserve. The case $d_2 = 1$ corresponds to a Markov chain.*

Since it is unable to capture the duration effects that are present, the Markov chain corresponding to $d_2 = 1$ performs, as anticipated, very badly. Maybe more surprisingly, the addition of just a single additional disability microstate corresponding to $d_2 = 2$ leads to significant improvements; this model may already be competitive, depending also on the trade-off between accuracy and computational load. Furthermore, and consistent with our expectations, the accuracy appears to further improve as the number of disability microstates, d_2, increases. Actually, any differences between the true and the aggregate model with $d_2 = 10$ might just be an expression of statistical noise.

3.A Proofs

This appendix contains the proofs of the results from Sections 3.4–3.5. In the following, we denote by $\boldsymbol{e}_{\widetilde{j}}$ the d_j-dimensional vector with a one in entry $\widetilde{j}$ and otherwise zeroes. To prove Proposition 3.4.1 and Theorem 3.4.3, we need the next lemma.

Lemma 3.A.1. *For $t \geq t_n$ and $k \neq y_n$ it holds that*

$$\mathbb{P}\big(t < T_{n+1} \leq t+h, \boldsymbol{X}(T_{n+1}) = \boldsymbol{k} \,\big|\, S_n = s_n\big) = \frac{\boldsymbol{\alpha}(s_n, t, k)\boldsymbol{e}_{\widetilde{k}}}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} h + o(h), \quad h \to 0,$$

where the d_{y_n}-dimensional row vector $\boldsymbol{\alpha}(s_n)$ is given by

$$\boldsymbol{\alpha}(s_n) = \boldsymbol{\pi}_1(0) \prod_{\ell=0}^{n-1} \prod_{t_\ell}^{t_{\ell+1}} (\boldsymbol{I} + \boldsymbol{M}_{y_\ell y_\ell}(x)\,\mathrm{d}x)\, \boldsymbol{M}_{y_\ell y_{\ell+1}}(t_{\ell+1}).$$

Proof. We give a proof by induction. First, we verify the identity for $n = 0$. Note that $\boldsymbol{\alpha}(0,1)\mathbf{1}_{d_1} = 1$ and that

$$\begin{aligned}
&\mathbb{P}\big(t < T_1 \leq t+h, \boldsymbol{X}(T_1) = \boldsymbol{k}\big) \\
&= \sum_{\widetilde{y}_0} \mathbb{P}\big(t < T_1 \leq t+h, \boldsymbol{X}(T_1) = \boldsymbol{k}, \boldsymbol{X}(t) = (1, \widetilde{y}_0)\big) \\
&= \sum_{\widetilde{y}_0} \bigg(\Big[o(h) + \mathbb{P}\big(T_1 \leq t+h, \boldsymbol{X}(t+h) = \boldsymbol{k} \,\big|\, t < T_1, \boldsymbol{X}(t) = (1, \widetilde{y}_0)\big)\Big] \\
&\qquad \mathbb{P}\big(t < T_1, \boldsymbol{X}(t) = (1, \widetilde{y}_0)\big)\bigg) \\
&= o(h) + h \sum_{\widetilde{y}_0} \mu_{(1,\widetilde{y}_0)\boldsymbol{k}}(t) \boldsymbol{\pi}_1(0) \prod_0^t (\boldsymbol{I} + \boldsymbol{M}_{11}(x)\,\mathrm{d}x)\, \boldsymbol{e}_{\widetilde{y}_0} \\
&= o(h) + h\, \boldsymbol{\pi}_1(0) \prod_0^t (\boldsymbol{I} + \boldsymbol{M}_{11}(x)\,\mathrm{d}x)\, \boldsymbol{M}_{1k}(t) \boldsymbol{e}_{\widetilde{k}} \\
&= o(h) + h\, \boldsymbol{\alpha}(s_1) \boldsymbol{e}_{\widetilde{k}}.
\end{aligned}$$

Collecting results confirms the identity for $n = 0$. Now suppose the identity holds for $n \in \mathbb{N}_0$. We want to establish the identity also for $n+1$. By assumption,

$$\mathbb{P}\big(t_{n+1} < T_{n+1} \leq t_{n+1} + \widetilde{h}, \boldsymbol{X}(T_{n+1}) = (y_{n+1}, \check{y}_{n+1}) \,\big|\, S_n = s_n\big)$$

$$= \frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})\boldsymbol{e}_{\check{y}_{n+1}}}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \widetilde{h} + o(\widetilde{h}).$$

In particular,

$$\begin{aligned} &\mathbb{P}\big(t_{n+1} < T_{n+1} \le t_{n+1} + \widetilde{h}, Y_{n+1} = y_{n+1} \,\big|\, S_n = s_n\big) \\ &= \frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})\mathbf{1}_{d_{y_{n+1}}}}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \widetilde{h} + o(\widetilde{h}). \end{aligned} \tag{3.A.1}$$

Furthermore,

$$\begin{aligned} &\mathbb{P}\big(t < T_{n+2}, \boldsymbol{X}(t) = \boldsymbol{y_{n+1}} \,\big|\, T_{n+1} = t_{n+1}, \boldsymbol{X}(t_{n+1}) = (y_{n+1}, \check{y}_{n+1})\big) \\ &= \boldsymbol{e}'_{\check{y}_{n+1}} \prod_{t_{n+1}}^{t} \big(\boldsymbol{I} + \boldsymbol{M}_{y_{n+1}y_{n+1}}(x)\,\mathrm{d}x\big)\, \boldsymbol{e}_{\widetilde{y}_{n+1}}, \end{aligned}$$

so that

$$\begin{aligned} &\mathbb{P}\big(t < T_{n+2}, \boldsymbol{X}(t) = \boldsymbol{y_{n+1}}, t_{n+1} < T_{n+1} \le t_{n+1} + \widetilde{h}, Y_{n+1} = y_{n+1} \,\big|\, S_n = s_n\big) \\ &= \frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_{n+1}}^{t} \big(\boldsymbol{I} + \boldsymbol{M}_{y_{n+1}y_{n+1}}(x)\,\mathrm{d}x\big)\, \boldsymbol{e}_{\widetilde{y}_{n+1}} \widetilde{h} + o(\widetilde{h}). \end{aligned}$$

Using this result, we find that

$$\begin{aligned} &\mathbb{P}\big(t < T_{n+2} \le t + h, \boldsymbol{X}(T_{n+2}) = \boldsymbol{k}, t_{n+1} < T_{n+1} \le t_{n+1} + \widetilde{h}, \\ &\quad Y_{n+1} = y_{n+1} \big| S_n = s_n\big) \\ &= \sum_{\widetilde{y}_{n+1}} \bigg(\Big[o(h) + \mathbb{P}\big(T_{n+2} \le t + h, \boldsymbol{X}(t+h) = \boldsymbol{k} \,\big|\, t < T_{n+2}, \boldsymbol{X}(t) = \boldsymbol{y_{n+1}}\big) \Big] \\ &\quad \mathbb{P}\big(t < T_{n+2}, \boldsymbol{X}(t) = \boldsymbol{y_{n+1}}, t_{n+1} < T_{n+1} \le t_{n+1} + \widetilde{h}, Y_{n+1} = y_{n+1} \,\big|\, S_n = s_n\big) \bigg) \\ &= \sum_{\widetilde{y}_{n+1}} \bigg(\Big[o(h) + \mu_{\boldsymbol{y_{n+1}}\boldsymbol{k}}(t) h \Big] \\ &\quad \Big[\frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_{n+1}}^{t} \big(\boldsymbol{I} + \boldsymbol{M}_{y_{n+1}y_{n+1}}(x)\,\mathrm{d}x\big)\, \boldsymbol{e}_{\widetilde{y}_{n+1}} \widetilde{h} + o(\widetilde{h}) \Big] \bigg), \end{aligned}$$

Combining this result with (3.A.1) allows us to conclude that

$$\begin{aligned} &\mathbb{P}\big(t < T_{n+2} \le t + h, \boldsymbol{X}(T_{n+2}) = \boldsymbol{k} \,\big|\, S_{n+1} = s_{n+1}\big) \\ &= o(h) + \frac{h \sum_{\widetilde{y}_{n+1}} \frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_{n+1}}^{t} \big(\boldsymbol{I} + \boldsymbol{M}_{y_{n+1}y_{n+1}}(x)\,\mathrm{d}x\big)\, \boldsymbol{e}_{\widetilde{y}_{n+1}} \mu_{\boldsymbol{y_{n+1}}\boldsymbol{k}}(t)}{\frac{\boldsymbol{\alpha}(s_n, t_{n+1}, y_{n+1})\mathbf{1}_{d_{y_{n+1}}}}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}}} \\ &= o(h) + h \frac{\boldsymbol{\alpha}(s_{n+1})}{\boldsymbol{\alpha}(s_{n+1})\mathbf{1}_{d_{y_{n+1}}}} \prod_{t_{n+1}}^{t} \big(\boldsymbol{I} + \boldsymbol{M}_{y_{n+1}y_{n+1}}(x)\,\mathrm{d}x\big)\, \boldsymbol{M}_{y_{n+1}k}(t) \boldsymbol{e}_{\widetilde{k}} \end{aligned}$$

$$= o(h) + \frac{\boldsymbol{\alpha}(s_{n+1}, t, k)\boldsymbol{e}_{\widetilde{k}}}{\boldsymbol{\alpha}(s_{n+1})\mathbf{1}_{d_{y_{n+1}}}} h.$$

This establishes the identity for $n+1$ and thus completes the proof. $\square$

Proof of Proposition 3.4.1. From Lemma 3.A.1 we immediately get that $\mathbb{P}(T_{n+1} \leq t, Y_{n+1} = k \mid S_n = s_n)$ is absolutely continuous with respect to the Lebesgue measure with density

$$\begin{aligned}
f^{(n+1)}(t, k|s_n) &= \frac{\boldsymbol{\alpha}(s_n, t, k)\mathbf{1}_{d_k}}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \\
&= \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_n k}(t)\mathbf{1}_{d_k}
\end{aligned}$$

for $t > t_n$ and $k \neq y_n$. In particular, $1 - \bar{F}^{(n+1)}(t|s_n)$ is absolutely continuous with respect to the Lebesgue measure with density

$$\begin{aligned}
f^{(n+1)}(t|s_n) &= \sum_{k \neq y_n} \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_n k}(t)\mathbf{1}_{d_k} \\
&= \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{m}_{y_n}(t)
\end{aligned} \tag{3.A.2}$$

for $t > t_n$; confer also with (3.3.2). Based on for instance the forward equations for product integrals, see Gill and Johansen (1990, Proposition 5 and 6), we may then argue that

$$\bar{F}^{(n+1)}(t|s_n) = \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \mathbf{1}_{d_{y_n}}, \quad t \geq t_n,$$

which proves the first assertion of the proposition. For the second part, we let $k \neq y_n$ and find that

$$\begin{aligned}
G^{(n+1)}(k|s_n, t_{n+1}) &= \frac{f^{(n+1)}(t_{n+1}, k|s_n)}{f^{(n+1)}(t_{n+1}|s_n)} \\
&= \frac{\frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_n k}(t_{n+1})\mathbf{1}_{d_k}}{\frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{m}_{y_n}(t_{n+1})} \\
&= \frac{\boldsymbol{\alpha}(s_n) \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{M}_{y_n k}(t_{n+1})\mathbf{1}_{d_k}}{\boldsymbol{\alpha}(s_n) \prod_{t_n}^{t_{n+1}} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{m}_{y_n}(t_{n+1})},
\end{aligned}$$

as desired. $\square$

Proof of Theorem 3.4.3. In the previous proof, we already noted that the distribution function $1 - \bar{F}^{(n+1)}(t|s_n)$ is absolutely continuous with respect to the Lebesgue measure with density

$$f^{(n+1)}(t|s_n) = \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{m}_{y_n}(t)$$

for $t > t_n$; see in particular (3.A.2). Then $\mathrm{d}\Lambda_{jk}(t) = \lambda_{jk}(t)\,\mathrm{d}t$ with

$$\lambda_{jk}(t) = \sum_{n \in \mathbb{N}_0} 1_{(T_n,\, T_{n+1}]}(t) 1_{(Y_n = j)} \frac{f^{(n+1)}(t|S_{n-1}, T_n, j)}{\bar{F}^{(n+1)}(t|S_{n-1}, T_n, j)} G^{(n+1)}(k|S_{n-1}, T_n, j, t),$$

confer with Jacobsen (2006, Proposition 4.4.1(b)(ii)). The result now follows from inserting the expressions for $\bar{F}^{(n+1)}$ and $G^{(n+1)}$ obtained in Proposition 3.4.1 along with the above expression for $f^{(n+1)}$. $\square$

Proof of Theorem 3.5.1. Due to the decomposition

$$\bar{p}_{\boldsymbol{j}}(t, s, z) = \sum_{n \in \mathbb{N}_0} 1_{[T_n,\, T_{n+1})}(t) \mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) > z \,|\, T_{n+1} > t, S_n),$$

it suffices to show that

$$\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) > z \,|\, T_{n+1} > t, S_n = s_n) = \widetilde{p}_{\boldsymbol{j}}(t, s, z; s_n),$$

whenever they are well-defined. The case $t_n \geq s - z$ is trivial, so suppose in the following that $t_n < s - z$. We find that

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) > z \,|\, T_{n+1} > t, S_n = s_n) \\
&= \sum_{\widetilde{y}_n} \frac{\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) > z, T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n} \,|\, S_n = s_n)}{\mathbb{P}(T_{n+1} > t \,|\, S_n = s_n)} \\
&= \sum_{\widetilde{y}_n} \Bigg(\frac{\mathbb{P}(T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n} \,|\, S_n = s_n)}{\mathbb{P}(T_{n+1} > t \,|\, S_n = s_n)} \\
&\qquad\qquad \mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{j}, U(s) > z \,|\, T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n}, S_n = s_n) \Bigg).
\end{aligned}$$

According to the first statement of Proposition 3.4.1,

$$\mathbb{P}(T_{n+1} > t \,|\, S_n = s_n) = \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \mathbf{1}_{d_{y_n}}.$$

Also, using similar techniques as in the proof of Proposition 3.4.1 and referring to Lemma 3.A.1, one may show that

$$\mathbb{P}(T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n} \,|\, S_n = s_n) = \frac{\boldsymbol{\alpha}(s_n)}{\boldsymbol{\alpha}(s_n)\mathbf{1}_{d_{y_n}}} \prod_{t_n}^{t} \left(\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x\right) \boldsymbol{e}_{\widetilde{y}_n}.$$

If $s - z \le t$, then

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,|\, T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n}, S_n = s_n) \\
&= 1_{(j=y_n)}\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, T_{n+1} > s \,|\, T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n}, S_n = s_n) \\
&= 1_{(j=y_n)}\boldsymbol{e}'_{\widetilde{y}_n} \prod_t^s (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\,\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}},
\end{aligned}$$

so that

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,|\, T_{n+1} > t, S_n = s_n) \\
&= 1_{(j=y_n)} \frac{\boldsymbol{\alpha}(s_n) \prod_{t_n}^s (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}}}{\boldsymbol{\alpha}(s_n) \prod_{t_n}^t (\boldsymbol{I} + \boldsymbol{M}_{y_n y_n}(x)\mathrm{d}x)\,\boldsymbol{1}_{d_{y_n}}},
\end{aligned}$$

which exactly equals $\widetilde{p}_{\boldsymbol{\jmath}}(t, s, z; s_n)$, confer with (3.5.10). If instead $s - z > t$, then the Markov property of $\boldsymbol{X}$ yields

$$\begin{aligned}
&\mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,|\, T_{n+1} > t, \boldsymbol{X}(t) = \boldsymbol{y_n}, S_n = s_n) \\
&= \sum_{\check{j}} \mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z, \boldsymbol{X}(s-z) = (j, \check{j}) \,|\, \boldsymbol{X}(t) = \boldsymbol{y_n}) \\
&= \sum_{\check{j}} \mathbb{P}(\boldsymbol{X}(s) = \boldsymbol{\jmath}, U(s) > z \,|\, \boldsymbol{X}(s-z) = (j, \check{j}))\mathbb{P}(\boldsymbol{X}(s-z) = (j, \check{j}) \,|\, \boldsymbol{X}(t) = \boldsymbol{y_n}) \\
&= \sum_{\check{j}} \boldsymbol{e}'_{\check{j}} \prod_{s-z}^s (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}} \boldsymbol{e}'_{\boldsymbol{y_n}} \prod_t^{s-z} (\boldsymbol{I} + \boldsymbol{M}(x)\mathrm{d}x)\,\boldsymbol{e}_{\boldsymbol{\jmath}} \\
&= \boldsymbol{e}'_{\boldsymbol{y_n}} \prod_t^{s-z} (\boldsymbol{I} + \boldsymbol{M}(x)\mathrm{d}x)\,\boldsymbol{E}_j \prod_{s-z}^s (\boldsymbol{I} + \boldsymbol{M}_{jj}(x)\mathrm{d}x)\,\boldsymbol{e}_{\widetilde{j}}.
\end{aligned}$$

Collecting results completes the proof. $\square$

Chapter 4

Aggregate Markov models in life insurance: estimation via the EM algorithm

This chapter is based on the manuscript *Ahmad and Bladt (2022a)*.

Abstract

In this paper, we consider statistical estimation of time-inhomogeneous aggregate Markov models. Unaggregated models, which corresponds to Markov chains, are commonly used in multi-state life insurance to model the biometric states of an insured. By aggregating microstates to each biometric state, we are able to model dependencies between transitions of the biometric states as well as the distribution of occupancy in these. This allows for non-Markovian modeling in general. Since only paths of the macrostates are observed, we develop an expectation-maximization (EM) algorithm to obtain maximum likelihood estimates of transition intensities on the micro level. Special attention is given to a semi-Markovian case, known as the reset property, which leads to simplified estimation procedures where EM algorithms for inhomogeneous phase-type distributions can be used as building blocks. We provide a numerical example of the latter in combination with piecewise constant transition rates in a three-state disability model with data simulated from a time-inhomogeneous semi-Markov model. Comparisons of our fits with more classic GLM-based fits as well as true and empirical distributions are provided to relate our model to existing models and their tools.

Keywords: Phase-type distributions; Parametric inference; EM algorithm; multi-state life insurance; semi-Markovianity

4.1 Introduction

In this paper, we consider statistical estimation of the finite state space, time-inhomogeneous aggregate Markov process introduced in the companion paper Ahmad, Bladt, and Furrer (2022). The term aggregate refers to certain states of the process being pooled into so-called macrostates. The interpretation of macrostates could for example be biometric or behavioral states in a life insurance context, like active, disabled, or free-policy. The pooled states are referred to as microstates and are introduced to improve the sojourn time distributions of the macrostates and introduce dependencies between transitions. This allows for non-Markovian modeling in general. Since only paths of the macrostates are observed, this corresponds to an incomplete data problem with respect to the underlying microstates, and we employ an expectation-maximization (EM) algorithm to obtain maximum likelihood estimates of transition intensities on the micro level.

The aggregate Markov model of Ahmad, Bladt, and Furrer (2022) may be considered as the underlying process of a time-inhomogeneous BMAP (Batch Markovian Arrival Processes, Latouche and Ramaswami, 1999), and contains as special cases time-homogeneous phase-type renewal processes (see Neuts, 1978), Markov modulated Poisson processes (see Rydén, 1994) and of course (time-homogeneous) BMAPs. In the aggregate Markov model, the sojourn time distributions are IPH and dependent, as shown in Ahmad, Bladt, and Furrer (2022, Proposition 4.1). Methods for fitting independent IPH distributions via the EM algorithm have been considered in Albrecher, Bladt, and Yslas (2022) for commuting sub-intensity matrix functions and in Ahmad, Bladt, and Bladt (2022) for general IPHs. Methods for the estimation of homogeneous BMAPs can be found in Breuer (2002, 2003).

In the multi-state life insurance context, however, we both need the time-inhomogeneity and dependency between transitions to properly capture age dependencies and duration effects. The nature of such models implies that sub-intensity matrices at different times may not commute, which is a crucial assumption in the independent IPH fitting of Albrecher, Bladt, and Yslas (2022). We, therefore, extend the general approach of Ahmad, Bladt, and Bladt (2022) to include the dependencies. This will provide the main contribution of the paper. Special attention is paid to the semi-Markovian case considered in Ahmad, Bladt, and Furrer (2022, Subsection 4.2), known as the reset property, where sojourn time distributions are IPH and independent. Here, we show how algorithms of Ahmad, Bladt, and Bladt (2022) partly can be used as inputs to our algorithms.

An important ingredient in our methods is the approximation of the models by piecewise constant transition rates. In general, transition probabilities of the Markov processes involved are solutions to ordinary differential equations of Kolmogorov type, the solution of which is denoted the product integral (see Gill and Johansen, 1990; Johansen, 1986). Assuming piecewise constant rates, the solutions can be expressed explicitly in terms of products of matrix exponentials. Furthermore, maximum likelihood estimation greatly simplifies and can be expressed in terms of multinomial and Poisson regressions based on a set of sufficient statistics in the different time intervals.

While we develop an EM algorithm for the general model, we only implement and apply it to data in the case where the reset property is satisfied along with piecewise constant transition rates; this relates to the approach in Ahmad, Bladt, and Bladt (2022). Here, we present a numerical example where macro data is simulated from a time-inhomogeneous semi-Markovian disability model commonly used in the context of disability insurance (see, e.g., Hoem, 1972; Helwich, 2008; Christiansen, 2012; Buchardt, Møller, and Schmidt, 2015). We compare our model fits with more classic GLM-based fits as well as true and empirical distributions to illustrate how the aggregate Markov model with the reset property is able to capture duration effects in these kinds of models.

The remainder of the paper is structured as follows. In Section 4.2 we set up the model and notation. Section 4.3 considers the estimation of completely observed aggregate Markov processes. Special attention is given to the piecewise constant case and the reset property, where links to Multinomial and Poisson regressions are provided. Then, in Section 4.4, an EM algorithm for fitting aggregate Markov models from observing only the macro process is developed. Special attention is devoted to the piecewise constant case and the reset property. The proof of the EM algorithm is deferred to Appendix 4.A. Finally, Section 4.5 contains a numerical example in a disability model.

4.2 The aggregate Markov model

We now present the aggregate Markov model introduced in the companion paper Ahmad, Bladt, and Furrer (2022), and some probabilistic properties of the model that are relevant to the present paper. Consider a jump process Z taking values on the finite set $\mathcal{J} = \{1, 2, ..., J\}$, $J \in \mathbb{N}$. We think of these as biometric or behavioral states governing the states of the insured in a life insurance context, for example, active, disabled, free-policy, or dead, and we thus refer to them as *macrostates*. To each macrostate $i \in \mathcal{J}$, a number $d_i \geq 1$ of *microstates* are assigned. The resulting state space is therefore

$$E = \{\boldsymbol{i} = (i, \tilde{i}) : i \in \mathcal{J}, \, \tilde{i} \in \{1, 2, ..., d_i\}\},$$

and the total number of microstates is $\bar{d} = \sum_{i \in \mathcal{J}} d_i$. Elements of E are in general denoted by bold letters such as $\boldsymbol{i} \in E$.

Now introduce a time-inhomogeneous Markov jump process

$$\boldsymbol{X} = \{\boldsymbol{X}(t)\}_{t\geq 0} = \{(X_1(t), X_2(t))\}_{t\geq 0}$$

on the state space E with transition intensity matrix function $\boldsymbol{M}(t)$. Then $X_1(t)$ keeps track of the macrostate, that is $Z(t) = X_1(t)$, while $X_2(t)$ identifies the current microstate contingent on the state of $X_1(t)$.

The transition intensity matrix function $\boldsymbol{M}(t)$ can be written on the following block form:

$$\boldsymbol{M}(t) = \begin{pmatrix} \boldsymbol{M}_{11}(t) & \boldsymbol{M}_{12}(t) & \cdots & \boldsymbol{M}_{1J}(t) \\ \boldsymbol{M}_{21}(t) & \boldsymbol{M}_{22}(t) & \cdots & \boldsymbol{M}_{2J}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{M}_{J1}(t) & \boldsymbol{M}_{J2}(t) & \cdots & \boldsymbol{M}_{JJ}(t) \end{pmatrix}, \tag{4.2.1}$$

where $\boldsymbol{M}_{ii}(t)$ are sub-intensity matrices of dimension $d_i \times d_i$ providing transition rates between the microstates of macrostate i at time t, and $\boldsymbol{M}_{ij}(t)$ are non-negative matrices of dimension $d_i \times d_j$ providing transition rates from microstates within macrostate i to microstates within macrostate j at time t. We denote an element of $\boldsymbol{M}(t)$ by $\mu_{\boldsymbol{\imath}\boldsymbol{\jmath}}(t)$, $\boldsymbol{\imath}, \boldsymbol{\jmath} \in E$. We assume that $Z(0) = X_1(0) \equiv 1$, so that the initial distribution of $\boldsymbol{X}$ is given by

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_1(0), \boldsymbol{0}),$$

where $\boldsymbol{\pi}_1(0)$ is the initial distribution of $X_2(0)$ among the microstates $1, 2, \ldots, d_1$. The transition (sub-)probability matrix functions within macrostates are given as the product integral (see Johansen, 1986; Gill and Johansen, 1990)

$$\bar{\boldsymbol{P}}_i(s,t) = \prod_s^t \left(\boldsymbol{I} + \boldsymbol{M}_{ii}(x)\,\mathrm{d}x\right), \quad i \in \mathcal{J}. \tag{4.2.2}$$

The vector $\boldsymbol{e}_{\tilde{i}}'\bar{\boldsymbol{P}}_i(s,t)$ then contains the distribution of $\boldsymbol{X}(t)$ within macrostate i on the event of staying in macrostate i in the whole time interval $[s,t]$, and given that $\boldsymbol{X}(s) = \boldsymbol{\imath}$; here and in the following, $\boldsymbol{e}_{\tilde{i}}$ denotes the d_i-dimensional column vector with one in entry $\tilde{i}$ and zeros otherwise, and primes denote matrix transposition. The matrix $\boldsymbol{M}_{ij}(t)\mathrm{d}t$, $j \neq i$, then contains the (infinitesimal) transition probabilities between microstates belonging to macrostates i and j, respectively. All in all, for some generic path of Z, represented via its associated marked point process $(T_i, Y_i)_{i\in\mathbb{N}_0}$ as $\mathcal{S}_n = (T_i, Y_i)_{i\leq n}$, the d_{y_n}-dimensional row vector (see Lemma A.1 in Ahmad, Bladt, and Furrer, 2022)

$$\boldsymbol{\alpha}(\mathfrak{s}_n) = \boldsymbol{\pi}_1(0) \prod_{\ell=0}^{n-1} \bar{\boldsymbol{P}}_{y_\ell}(t_\ell, t_{\ell+1}) \boldsymbol{M}_{y_\ell y_{\ell+1}}(t_{\ell+1}) \tag{4.2.3}$$

provides the (defective) distribution in macrostate y_n at time t_n. In particular, according to Ahmad, Bladt, and Furrer (2022, Remark 4.2), the sojourn times are inhomogeneous phase-type distributed (IPH, Albrecher and Bladt, 2019) and dependent on past jump times and transitions:

$$T_{n+1} - T_n \,|\, \mathcal{S}_n \sim \mathrm{IPH}\bigg(\frac{\boldsymbol{\alpha}(\mathcal{S}_n)}{\boldsymbol{\alpha}(\mathcal{S}_n)\mathbf{1}_{d_{y_n}}},\, \boldsymbol{M}_{Y_n Y_n}(T_n + \bullet)\bigg). \tag{4.2.4}$$

The corresponding exit rate function out of macrostate i is then given as the column vector function

$$\boldsymbol{m}_i(t) = -\boldsymbol{M}_{ii}(t)\mathbf{1}_{d_i} = \sum_{\substack{j\in\mathcal{J}\\ j\neq i}} \boldsymbol{M}_{ij}(t)\mathbf{1}_{d_j}. \tag{4.2.5}$$

Throughout the paper, we pay special attention to the case where the *reset property* introduced in Ahmad, Bladt, and Furrer (2022, (3.3)) is satisfied. Here $\boldsymbol{M}_{ij}(t)$, $j \neq i$, is a matrix of rank one on the form

$$\boldsymbol{M}_{ij}(t) = \boldsymbol{\beta}_{ij}(t)\boldsymbol{\pi}_j(t), \tag{4.2.6}$$

where $\boldsymbol{\beta}_{ij}(t)$ is a d_i-dimensional non-negative column vector function and $\boldsymbol{\pi}_j(t)$ is a d_j-dimensional non-negative row vector function with $\boldsymbol{\pi}_j(t)\mathbf{1}_{d_j} = 1$. In this case,

$$\boldsymbol{m}_i(t) = \sum_{\substack{j\in\mathcal{J}\\ j\neq i}} \boldsymbol{\beta}_{ij}(t), \tag{4.2.7}$$

and, according to Ahmad, Bladt, and Furrer (2022, Remark 4.5), the conditional sojourn time distributions (4.2.4) become independent of past jump times and transitions:

$$T_{n+1} - T_n \,|\, \mathcal{S}_n \sim \mathrm{IPH}(\boldsymbol{\pi}_{Y_n}(T_n),\, \boldsymbol{M}_{Y_n Y_n}(T_n + \bullet)). \tag{4.2.8}$$

These simplifications imply a specific time-inhomogeneous semi-Markovian structure to the macrostate process Z, cf. Ahmad, Bladt, and Furrer (2022, Subsection 4.2), which explains the focus on these type of models in our numerical example in Section 4.5.

In this paper, we develop methods for statistical fitting of the aggregate Markov model, namely estimation of the micro intensities $\mu_{\boldsymbol{ij}}$ based on independent observations of the trajectories of the macrostate process Z. Since this leads to incomplete data with respect to the underlying macro-micro state process $\boldsymbol{X}$, we employ an expectation-maximization (EM) algorithm to obtain maximum likelihood estimations of the micro intensities. We develop a general EM algorithm and implement it in the case where the reset property (4.2.6) is satisfied along with piecewise constant transition rates.

4.2.1 Piecewise constant transition rates

Following Ahmad, Bladt, and Bladt (2022, Section 2.1), suppose that the transition rates are piecewise constant on a grid $s_0 = 0 < s_1 < \cdots < s_{K-1} < \infty = s_K$ of K

time points, $K \in \mathbb{N}$, with values

$$\boldsymbol{M}(s) = \boldsymbol{M}^k = \left\{\mu^k_{\boldsymbol{i}\boldsymbol{j}}\right\}_{\boldsymbol{i},\boldsymbol{j}\in E}, \quad s \in (s_{k-1}, s_k], \quad k \in \{1, \dots, K\}. \tag{4.2.9}$$

Introducing $k(x)$ as the unique $k \in \{1, \dots, K\}$ satisfying that $x \in (s_{k-1}, s_k]$, we now have that the transition (sub-)probability matrix functions within macrostates (4.2.2) simplify to a product of matrix exponentials on the form

$$\bar{\boldsymbol{P}}_i(s,t) = \mathrm{e}^{\boldsymbol{M}^{k(s)}_{ii}\left(s_{k(s)}-s\right)} \left(\prod_{\ell=k(s)+1}^{k(t)-1} \mathrm{e}^{\boldsymbol{M}^{\ell}_{ii}(s_\ell - s_{\ell-1})} \right) \mathrm{e}^{\boldsymbol{M}^{k(t)}_{ii}\left(t-s_{k(t)-1}\right)}, \tag{4.2.10}$$

with the convention that empty product integrals equals the identity matrix. The defective distribution (4.2.3) at time t_n then also simplify, to

$$\boldsymbol{\alpha}(\mathfrak{z}_n) = \boldsymbol{\pi}_1(0) \prod_{\ell=0}^{n-1} \bar{\boldsymbol{P}}_{y_\ell}(t_\ell, t_{\ell+1}) \boldsymbol{M}^{k(t_{\ell+1})}_{y_\ell y_{\ell+1}}, \tag{4.2.11}$$

with the matrices $\bar{\boldsymbol{P}}_{y_\ell}(t_\ell, t_{\ell+1})$ being on the form (4.2.10). We give special attention to this case being satisfied along with the reset property when we develop our algorithms in this paper, as they will provide simplifications similar to those in Ahmad, Bladt, and Bladt (2022).

Here it may be noted that if the resulting exit rates (4.2.5) are different between two sub-intervals, the density of the conditional sojourn time distributions (4.2.4) become discontinuous at the corresponding grid point between the two sub-intervals. This follows by similar arguments as those made in Ahmad, Bladt, and Bladt (2022, Subsection 2.1).

4.3 The case of complete micro data

We now consider the complete data case where trajectories of the underlying macro-micro state process $\boldsymbol{X}$ is fully observed, which corresponds to methods known from inference of time-inhomogeneous Markov jump processes on finite state spaces; we refer to Andersen et al. (1993) for a detailed exposition on this. The approach and notation of this section largely follow that of Ahmad, Bladt, and Bladt (2022, Section 3.1-3.2).

4.3.1 General case

Suppose that we observe $N \in \mathbb{N}$ i.i.d. realizations of the Markov jump process $\boldsymbol{X}$ of macro-micro states on some time interval $[0, T]$, where $T > 0$ is a given and fixed time horizon; represent the (fictive) data by $\boldsymbol{X} = (\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)})$. Denote with $\boldsymbol{N} = (N^{(1)}, \dots, N^{(N)})$ the corresponding data of the multivariate counting process associated to $\boldsymbol{X}$, that is, $N^{(n)}$, $n = 1, \dots, N$, have components

$$N^{(n)}_{\boldsymbol{i}\boldsymbol{j}}(t) = \#\left\{s \in (0,t] : \boldsymbol{X}^{(n)}(s-) = \boldsymbol{i},\ \boldsymbol{X}^{(n)}(s) = \boldsymbol{j}\right\}.$$

Parametrizing the transition rates on the micro level with a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is some finite-dimensional, suitably regular parameter space with non-empty interior, such that

$$\boldsymbol{M}(s) = \boldsymbol{M}(s;\boldsymbol{\theta}),$$

we have that the likelihood function for the joint parameter $(\boldsymbol{\pi}_1, \boldsymbol{\theta})$ is given by

$$\begin{aligned}
\mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) &= \mathcal{L}_0^{\boldsymbol{X}}(\boldsymbol{\pi}_1) \prod_{\substack{\boldsymbol{i},\boldsymbol{j} \in E \\ \boldsymbol{j} \neq \boldsymbol{i}}} \mathcal{L}_{\boldsymbol{ij}}^{\boldsymbol{X}}(\boldsymbol{\theta}), \\
\mathcal{L}_0^{\boldsymbol{X}}(\boldsymbol{\pi}_1) &= \prod_{r=1}^{d_1} \pi_{(1,r)}(0)^{B_{(1,r)}(0)}, \qquad (4.3.1) \\
\mathcal{L}_{\boldsymbol{ij}}^{\boldsymbol{X}}(\boldsymbol{\theta}) &= \exp\left(\int_{(0,T]} \log\big(\mu_{\boldsymbol{ij}}(s;\boldsymbol{\theta})\big) \mathrm{d}N_{\boldsymbol{ij}}(s) - \int_0^T I_{\boldsymbol{i}}(s) \mu_{\boldsymbol{ij}}(s;\boldsymbol{\theta}) \mathrm{d}s \right),
\end{aligned}$$

where, for $\boldsymbol{i}, \boldsymbol{j} \in E$, $\boldsymbol{j} \neq \boldsymbol{i}$, and $s \in [0,T]$,

$$I_{\boldsymbol{i}}(s) = \sum_{n=1}^{N} \mathbb{1}_{(\boldsymbol{X}^{(n)}(s) = \boldsymbol{i})} \qquad \text{and} \qquad N_{\boldsymbol{ij}}(s) = \sum_{n=1}^{N} N_{\boldsymbol{ij}}^{(n)}(s), \qquad (4.3.2)$$

with $B_{(1,r)}(0) = I_{(1,r)}(0)$.

Then $I_{\boldsymbol{i}}(s)$ gives the number of observations in state $\boldsymbol{i}$ at time s, while $N_{\boldsymbol{ij}}(s)$ gives the total number of jumps observed from state $\boldsymbol{i}$ to $\boldsymbol{j}$ on $[0,s]$. In particular, $B_{(1,r)}(0)$ then becomes the total number of initiations in microstate r of macrostate 1 observed. When we later consider the case where the reset property is satisfied along with piecewise constant transition rates, we encounter more kinds of initiations in macrostates that adds to this notion.

The corresponding log-likelihood $L^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) = \log \mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta})$ is given by

$$\begin{aligned}
L^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) &= L_0^{\boldsymbol{X}}(\boldsymbol{\pi}_1) + \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in E \\ \boldsymbol{j} \neq \boldsymbol{i}}} L_{\boldsymbol{ij}}^{\boldsymbol{X}}(\boldsymbol{\theta}), \qquad (4.3.3) \\
L_0^{\boldsymbol{X}}(\boldsymbol{\pi}_1) &= \sum_{r=1}^{d_1} B_{(1,r)}(0) \log(\pi_{(1,r)}(0)), \\
L_{\boldsymbol{ij}}^{\boldsymbol{X}}(\boldsymbol{\theta}) &= \int_{(0,T]} \log\big(\mu_{\boldsymbol{ij}}(s;\boldsymbol{\theta})\big) \mathrm{d}N_{\boldsymbol{ij}}(s) - \int_0^T I_{\boldsymbol{i}}(s) \mu_{\boldsymbol{ij}}(s;\boldsymbol{\theta}) \mathrm{d}s, \qquad (4.3.4)
\end{aligned}$$

and the MLE of $(\boldsymbol{\pi}_1, \boldsymbol{\theta})$ is then found by maximizing the log-likelihood:

$$(\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\theta}}) = \underset{(\boldsymbol{\pi}_1,\, \boldsymbol{\theta})}{\arg\max}\; L^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta}).$$

The product structure of the likelihood (4.3.1) in $\boldsymbol{\pi}_1$ and $\boldsymbol{\theta}$ gives that we can estimate these separately via their respective likelihoods $\mathcal{L}_0^{\boldsymbol{X}}$ and $\mathcal{L}_{\boldsymbol{ij}}^{\boldsymbol{X}}$, $\boldsymbol{i}, \boldsymbol{j} \in E$, $\boldsymbol{j} \neq \boldsymbol{i}$. For

$\boldsymbol{\pi}_1$, one realizes that the likelihood $\mathcal{L}_0^{\boldsymbol{X}}$ is proportional to the likelihood obtained from viewing $(B_{(1,1)}(0), \ldots, B_{(1,d_1)}(0))$ as an observation from the Multinomial$(N, \boldsymbol{\pi}_1(0))$ distribution, where N is considered fixed. Hence, the MLE of $\boldsymbol{\pi}_1$ is explicitly given by

$$\hat{\pi}_{(1,r)}(0) = \frac{B_{(1,r)}(0)}{N}. \tag{4.3.5}$$

The MLE of $\boldsymbol{\theta}$ is then given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in E \\ \boldsymbol{j} \neq \boldsymbol{i}}} L_{\boldsymbol{i}\boldsymbol{j}}^{\boldsymbol{X}}(\boldsymbol{\theta}),$$

which, in general, requires numerical methods for optimization. Similar discussions are seen in, e.g., Ahmad, Bladt, and Bladt (2022, Section 3.1).

4.3.2 Reset property

We now assume the reset property (4.2.6) is satisfied. The setup remains that of Subsection 4.3.1, except that we now, due to the nature of the exit rates $\boldsymbol{\beta}_{ij}$ and initial distributions $\boldsymbol{\pi}_j$ playing distinct roles, extend the parameter space to $\boldsymbol{\Theta} \times \boldsymbol{H}$, such that, for $(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \boldsymbol{\Theta} \times \boldsymbol{H}$,

$$\boldsymbol{M}_{ij}(s; \boldsymbol{\theta}, \boldsymbol{\eta}) = \boldsymbol{\beta}_{ij}(s; \boldsymbol{\theta}) \boldsymbol{\pi}_j(s; \boldsymbol{\eta}), \qquad j \neq i. \tag{4.3.6}$$

This parameterization allows for separate estimations of exit rates and initial distributions within the reset property. Note that we implicitly also set $\boldsymbol{\pi}_1(0) = \boldsymbol{\pi}_1(0; \boldsymbol{\eta})$, so that we allow for the possibility of $\boldsymbol{\pi}_1(0)$ to be regressed against the other initial distributions at the different time points.

Having this setup, we see by splitting the likelihood contributions for the different transitions, $\mathcal{L}_{\boldsymbol{i}\boldsymbol{j}}^{\boldsymbol{X}}$, between those within macrostates and those between macrostates that the likelihood (4.3.1) now simplifies to, using that $\boldsymbol{\pi}_j(s; \boldsymbol{\eta}) \mathbf{1}_{d_j} = 1$ for all j,

$$\mathcal{L}^{\boldsymbol{X}}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \prod_{\boldsymbol{i} \in E} \mathcal{L}_{\boldsymbol{i}}^{\boldsymbol{X}}(\boldsymbol{\eta}) \prod_{\substack{\check{i}=1 \\ \check{i} \neq \tilde{i}}}^{d_i} \mathcal{L}_{\boldsymbol{i}(i,\check{i})}^{\boldsymbol{X}}(\boldsymbol{\theta}) \prod_{\substack{j \in \mathcal{J} \\ j \neq i}} \mathcal{L}_{\boldsymbol{i}j}^{\boldsymbol{X}}(\boldsymbol{\theta}),$$

with $\mathcal{L}_{\boldsymbol{i}(i,\check{i})}^{\boldsymbol{X}}$ as in (4.3.1), and

$$\mathcal{L}_{\boldsymbol{i}}^{\boldsymbol{X}}(\boldsymbol{\eta}) = \exp\bigg(\int_{[0,T]} \log\left(\pi_{\boldsymbol{i}}(s; \boldsymbol{\eta})\right) \mathrm{d}N_{\boldsymbol{i}}(s) \bigg), \tag{4.3.7}$$

$$\mathcal{L}_{\boldsymbol{i}j}^{\boldsymbol{X}}(\boldsymbol{\theta}) = \exp\bigg(\int_{(0,T]} \log\left(\beta_{\boldsymbol{i}j}(s; \boldsymbol{\theta})\right) \mathrm{d}N_{\boldsymbol{i}j}(s) - \int_0^T I_{\boldsymbol{i}}(s) \beta_{\boldsymbol{i}j}(s; \boldsymbol{\theta}) \, \mathrm{d}s \bigg), \tag{4.3.8}$$

where we define $N_{\boldsymbol{i}j}$ and $N_{\boldsymbol{i}}$ as the aggregated processes

$$N_{\boldsymbol{i}j}(s) = \sum_{\tilde{j}=1}^{d_j} N_{\boldsymbol{i}\boldsymbol{j}}(s) \quad \text{and} \quad N_{\boldsymbol{i}}(s) = \begin{cases} \sum_{\substack{\boldsymbol{j}\in E \\ j\neq i}} N_{\boldsymbol{j}\boldsymbol{i}}(s) & \text{for } s > 0, \\ B_{\boldsymbol{i}}(0) & \text{for } s = 0,\ i = 1, \\ 0 & \text{Otherwise.} \end{cases} \tag{4.3.9}$$

Note that we rather untraditionally, but for notational convenience, couple the number of initiations at time 0 with the counting process counting the number of jumps into a macrostate in our definition of $N_{\boldsymbol{i}}$. This is related to the aforementioned possibility of regressing the initial distribution $\boldsymbol{\pi}_1(0;\boldsymbol{\eta})$ at time 0 with the other initial distributions $\boldsymbol{\pi}_j(\cdot;\boldsymbol{\eta})$, where this definition allows us to unify computations.

The corresponding log-likelihood (4.3.3) takes the form

$$L^{\boldsymbol{X}}(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{\boldsymbol{i}\in E} \left(L^{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{\eta}) + \sum_{\substack{\check{i}=1 \\ \check{i}\neq\tilde{i}}}^{d_i} L^{\boldsymbol{X}}_{\boldsymbol{i}(i,\check{i})}(\boldsymbol{\theta}) + \sum_{\substack{j\in\mathcal{J} \\ j\neq i}} L^{\boldsymbol{X}}_{\boldsymbol{i}j}(\boldsymbol{\theta}) \right), \tag{4.3.10}$$

with $L^{\boldsymbol{X}}_{\boldsymbol{i}(i,\check{i})}$ as in (4.3.4), and

$$\begin{aligned} L^{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{\eta}) &= \int_{[0,T]} \log\left(\pi_{\boldsymbol{i}}(s;\boldsymbol{\eta})\right) \mathrm{d}N_{\boldsymbol{i}}(s), \\ L^{\boldsymbol{X}}_{\boldsymbol{i}j}(\boldsymbol{\theta}) &= \int_{(0,T]} \log\left(\beta_{\boldsymbol{i}j}(s;\boldsymbol{\theta})\right) \mathrm{d}N_{\boldsymbol{i}j}(s) - \int_0^T I_{\boldsymbol{i}}(s)\beta_{\boldsymbol{i}j}(s;\boldsymbol{\theta})\,\mathrm{d}s. \end{aligned} \tag{4.3.11}$$

The MLE of $(\boldsymbol{\theta},\boldsymbol{\eta})$ is then found by maximizing the relevant log-likelihood contributions:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{\boldsymbol{i}\in E} \left(\sum_{\substack{\check{i}=1 \\ \check{i}\neq\tilde{i}}}^{d_i} L^{\boldsymbol{X}}_{\boldsymbol{i}(i,\check{i})}(\boldsymbol{\theta}) + \sum_{\substack{j\in\mathcal{J} \\ j\neq i}} L^{\boldsymbol{X}}_{\boldsymbol{i}j}(\boldsymbol{\theta}) \right),$$

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \sum_{\boldsymbol{i}\in E} L^{\boldsymbol{X}}_{\boldsymbol{i}}(\boldsymbol{\eta}),$$

which also here, in general, requires numerical methods for optimization.

4.3.3 Piecewise constant transition rates

Consider again the general case of Subsection 4.3.1, and assume now that the transition intensity matrix function $\boldsymbol{M}(\cdot;\boldsymbol{\theta})$ is piecewise constant on the form (4.2.9). Then the likelihood contributions for the transitions between states, $\mathcal{L}^{\boldsymbol{X}}_{\boldsymbol{i}\boldsymbol{j}}$, simplify to:

$$\mathcal{L}^{\boldsymbol{X}}_{\boldsymbol{i}\boldsymbol{j}}(\boldsymbol{\theta}) = \prod_{k=1}^{K} \left(\mu^k_{\boldsymbol{i}\boldsymbol{j}}(\boldsymbol{\theta})\right)^{O_{\boldsymbol{i}\boldsymbol{j}}(k)} \exp\left(-\mu^k_{\boldsymbol{i}\boldsymbol{j}}(\boldsymbol{\theta})E_{\boldsymbol{i}}(k)\right), \tag{4.3.12}$$

where $O_{\boldsymbol{i}\boldsymbol{j}}(k)$ is the total number of *occurrences* of transitions from state $\boldsymbol{i}$ to $\boldsymbol{j}$ in the time interval $(s_{k-1}, s_k]$, and $E_{\boldsymbol{i}}(k)$ is the total time spent in state $\boldsymbol{i}$ in the time interval $(s_{k-1}, s_k]$, the so-called *exposure*, given by

$$O_{\boldsymbol{i}\boldsymbol{j}}(k) = \int_{(s_{k-1}, s_k]} \mathrm{d}N_{\boldsymbol{i}\boldsymbol{j}}(t) \qquad \text{and} \qquad E_{\boldsymbol{i}}(k) = \int_{s_{k-1}}^{s_k} I_{\boldsymbol{i}}(t)\,\mathrm{d}t. \tag{4.3.13}$$

The corresponding log-likelihood contributions take the form

$$L_{\boldsymbol{i}\boldsymbol{j}}^{\boldsymbol{X}}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \Big(O_{\boldsymbol{i}\boldsymbol{j}}(k)\log\big(\mu_{\boldsymbol{i}\boldsymbol{j}}^{k}(\boldsymbol{\theta})\big) - \mu_{\boldsymbol{i}\boldsymbol{j}}^{k}(\boldsymbol{\theta})E_{\boldsymbol{i}}(k)\Big). \tag{4.3.14}$$

Thus, in the case of piecewise constant transition rates, the occurrences and exposures in the different time intervals, along with the number of initiations in the different microstates of macrostate 1,

$$\big\{\big(B_{(1,r)}(0),\, O_{\boldsymbol{i}\boldsymbol{j}}(k),\, E_{\boldsymbol{i}}(k)\big)\big\}_{k\in\{1,\ldots,K\},\, r\in\{1,\ldots,d_1\},\, \boldsymbol{i},\boldsymbol{j}\in E,\, \boldsymbol{j}\neq\boldsymbol{i}},$$

are sufficient statistics. One even notes that the resulting likelihood, (4.3.1) combined with (4.3.12), is proportional to the likelihood obtained from independent observations

$$\begin{aligned} &\big(B_{(1,1)}(0), \ldots, B_{(1,d_1)}(0)\big), \\ &\big(O_{\boldsymbol{i}\boldsymbol{j}}(k), \quad k \in \{1,\ldots,K\}, \quad \boldsymbol{i},\boldsymbol{j}\in E,\ \boldsymbol{j}\neq\boldsymbol{i}\big), \end{aligned} \tag{4.3.15}$$

where

$$\begin{aligned} \big(B_{(1,1)}(0), \ldots, B_{(1,d_1)}(0)\big) \quad &\text{is} \quad \text{Multinomial}(N, \boldsymbol{\pi}_1(0)) - \text{distributed}, \\ O_{\boldsymbol{i}\boldsymbol{j}}(k) \quad &\text{is} \quad \text{Poisson}\big(E_{\boldsymbol{i}}(k)\mu_{\boldsymbol{i}\boldsymbol{j}}^{k}(\boldsymbol{\theta})\big) - \text{distributed}, \end{aligned} \tag{4.3.16}$$

with N and $E_{\boldsymbol{i}}(k)$ considered fixed. Hence, the MLE of $\boldsymbol{\pi}_1(0)$ remains explicitly given by (4.3.5), while the MLE of $\boldsymbol{\theta}$ can be obtained from Poisson regressions of the occurrences against the times on the grid, which can be carried using standard software packages. For example, if the intensities $\mu_{\boldsymbol{i}\boldsymbol{j}}^{k}(\boldsymbol{\theta})$ are exponential functions of $\boldsymbol{\theta}$, a Poisson regression with log-link function and log-exposure as offset can be carried out, corresponding to fitting the model

$$\log(\mu_{\boldsymbol{i}\boldsymbol{j}}(s;\boldsymbol{\theta})) = \sum_{r=1}^{q} \theta_{\boldsymbol{i}\boldsymbol{j}}^{(r)} g^{(r)}(s), \tag{4.3.17}$$

for suitably regular known functions $g^{(r)}$, with a common choice being $g^{(r)}(s) = s^r$. The predictions at s_k then constitutes the MLEs of the intensities.

In the special case where each of the parameters in $\boldsymbol{\theta}$ are the transition rates in the different time intervals, that is, $\boldsymbol{\theta} = \Big(\theta_{\boldsymbol{i}\boldsymbol{j}}^{k}\Big)_{\boldsymbol{i},\boldsymbol{j}\in E,\, \boldsymbol{j}\neq\boldsymbol{i},\, k=1,\ldots,K}$ such that

$$\mu_{\boldsymbol{i}\boldsymbol{j}}^{k}(\boldsymbol{\theta}) = \theta_{\boldsymbol{i}\boldsymbol{j}}^{k},$$

the MLE of $\boldsymbol{\theta}$ simplify to so-called occurrence-exposure rates (cf. also Asmussen, Nerman, and Olsson, 1996; Ahmad, Bladt, and Bladt, 2022):

$$\hat{\theta}^k_{\tilde{\imath}\tilde{\jmath}} = \frac{O_{\tilde{\imath}\tilde{\jmath}}(k)}{E_{\tilde{\imath}}(k)}.$$

This can be seen as a direct "non-parametric" approach to estimate the micro intensities in the different time intervals, which then is a special case of the general parametric approach. The assumption of piecewise constant transition rates is often seen as an approximation to continuous versions obtained when the number of grid points tends to infinity. The observations of this subsection largely follow the observations made in Aalen, Borgan, and Gjessing (2008, Section 5).

4.3.4 Reset property with piecewise constant transition rates

We now assume the reset property (4.2.6) in combination with piecewise constant transition rates on the form (4.2.9), so that for $j \neq i$, $k \in \{1, \ldots, K\}$, and $s \in (s_{k-1}, s_k]$,

$$\begin{aligned} \boldsymbol{\beta}_{ij}(s; \boldsymbol{\theta}) &= \boldsymbol{\beta}^k_{ij}(\boldsymbol{\theta}), \\ \boldsymbol{\pi}_j(s; \boldsymbol{\eta}) &= \boldsymbol{\pi}^k_j(\boldsymbol{\eta}), \end{aligned}$$

with $\boldsymbol{\pi}_1(0; \boldsymbol{\eta}) = \boldsymbol{\pi}^0_1(\boldsymbol{\eta})$. The transition rates between macrostates are then on the form

$$\boldsymbol{M}^k_{ij}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \boldsymbol{\beta}^k_{ij}(\boldsymbol{\theta}) \boldsymbol{\pi}^k_j(\boldsymbol{\eta}). \tag{4.3.18}$$

In this case, the likelihood contributions for transitions between macrostates (4.3.7)-(4.3.8) simplify to

$$\begin{aligned} \mathcal{L}^{\boldsymbol{X}}_{\tilde{\imath}}(\boldsymbol{\eta}) &= \prod_{k=0}^{K} \pi^k_{\tilde{\imath}}(\boldsymbol{\eta})^{B_{\tilde{\imath}}(k)}, \\ \mathcal{L}^{\boldsymbol{X}}_{\tilde{\imath}j}(\boldsymbol{\theta}) &= \prod_{k=1}^{K} \beta^k_{\tilde{\imath}j}(\boldsymbol{\theta})^{O_{\tilde{\imath}j}(k)} \exp\big(-\beta^k_{\tilde{\imath}j}(\boldsymbol{\theta}) E_{\tilde{\imath}}(k)\big), \end{aligned} \tag{4.3.19}$$

where, for $k \in \{1, \ldots, K\}$, $B_{\tilde{\imath}}(k)$ is the total number of initiations in microstate $\tilde{\imath}$ in the time interval $(s_{k-1}, s_k]$ resulting from jumps into macrostate i, and $O_{\tilde{\imath}j}(k)$ is the total number of transitions in time interval $(s_{k-1}, s_k]$ from macrostate i to j happening from microstate $\tilde{\imath}$:

$$B_{\tilde{\imath}}(k) = \sum_{\substack{\tilde{\jmath} \in E \\ j \neq i}} O_{\tilde{\jmath}\tilde{\imath}}(k) \qquad \text{and} \qquad O_{\tilde{\imath}j}(k) = \sum_{\tilde{j}=1}^{d_j} O_{\tilde{\imath}\tilde{\jmath}}(k). \tag{4.3.20}$$

The corresponding log-likelihood contributions simplify to

$$L_{\boldsymbol{i}}^{\boldsymbol{X}}(\boldsymbol{\eta}) = \sum_{k=0}^{K} B_{\boldsymbol{i}}(k) \log\bigl(\pi_{\boldsymbol{i}}^{k}(\boldsymbol{\eta})\bigr), \tag{4.3.21}$$

$$L_{\boldsymbol{i}j}^{\boldsymbol{X}}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \bigl(O_{\boldsymbol{i}j}(k) \log\bigl(\beta_{\boldsymbol{i}j}^{k}(\boldsymbol{\theta})\bigr) - \beta_{\boldsymbol{i}j}^{k}(\boldsymbol{\theta}) E_{\boldsymbol{i}}(k)\bigr). \tag{4.3.22}$$

Consequently, in the case where the reset property is satisfied in combination with piecewise constant transition rates, (4.2.9) and (4.3.18), the sufficient statistics regarding the occurrences between macrostates reduce to those of (4.3.20). In fact, by inserting the simplified likelihood contributions (4.3.19) into the general piecewise constant case (4.3.12), which again are inserted into the general likelihood (4.3.1), we realize that it now is proportional to the likelihood obtained from independent observations

$$\bigl(B_{\boldsymbol{i}}(k), \qquad k = 0, \dots, K,\ \boldsymbol{i} \in E\bigr),$$

$$\Bigl(\bigl(O_{\boldsymbol{i}(i,\check{i})}(k), O_{\boldsymbol{i}j}(k)\bigr),\ k = 1, \dots, K,\ \boldsymbol{i} \in E,\ \check{i} \in \{1, \dots, d_i\},\ \check{i} \neq \tilde{i},\ j \in \mathcal{J},\ j \neq i\Bigr),$$

where

$$\begin{aligned} \bigl(B_{(i,1)}(k), \dots, B_{(i,d_i)}(k)\bigr) \quad &\text{is} \quad \text{Multinomial}\bigl(B_i(k), \boldsymbol{\pi}_i^k(\boldsymbol{\eta})\bigr) - \text{distributed}, \\ O_{\boldsymbol{i}(i,\check{i})}(k) \quad &\text{is} \quad \text{Poisson}\Bigl(E_{\boldsymbol{i}}(k)\mu_{\boldsymbol{i}(i,\check{i})}^{k}(\boldsymbol{\theta})\Bigr) - \text{distributed}, \\ O_{\boldsymbol{i}j}(k) \quad &\text{is} \quad \text{Poisson}\bigl(E_{\boldsymbol{i}}(k)\beta_{\boldsymbol{i}j}^{k}(\boldsymbol{\theta})\bigr) - \text{distributed}, \end{aligned}$$

with $B_i(k) = \sum_{\tilde{i}=1}^{d_i} B_{\boldsymbol{i}}(k)$, $k \in \{1, \dots, K\}$, being the total number of jumps to macrostate i observed in $(s_{k-1}, s_k]$; for $k = 0$ and $i = 1$, we have $B_1(0) = N$, cf. also (4.3.16). Hence, the MLE of $\boldsymbol{\theta}$ is obtained from similar kinds of Poisson regressions as those in Subsection 4.3.3, but the MLE of $\boldsymbol{\eta}$ can now be obtained from multinomial regressions of the number of initiations against the times on the grid, which also can be carried using standard software packages.

For example, if for a fixed macrostate $i \in \mathcal{J}$, the probabilities $\pi_{\boldsymbol{i}}^{k}(\boldsymbol{\eta})$ are exponential functions of $\boldsymbol{\eta}$ (relative to the probability $\pi_{(i,d_i)}^{k}(\boldsymbol{\eta})$ in the last microstate d_i, say), then a multinomial logistic regression can be carried out, corresponding to fitting the model

$$\pi_{\boldsymbol{i}}(s;\boldsymbol{\eta}) = \frac{\exp\Bigl(\mathbb{1}_{(\tilde{i} \neq d_i)} \sum_{r=1}^{q} \eta_{\boldsymbol{i}}^{(r)} g^{(r)}(s)\Bigr)}{1 + \sum_{\check{i}=1}^{d_i - 1} \exp\Bigl(\sum_{r=1}^{q} \eta_{(i,\check{i})}^{(r)} g^{(r)}(s)\Bigr)}, \tag{4.3.23}$$

where the functions $g^{(r)}$ are as in (4.3.17). The predictions at s_k then constitute the MLEs of the initial distributions. Similar types of multinomial logistic regressions for initial distributions of (inhomogeneous) phase-type distributions are performed in Bladt and Yslas (2022) and Albrecher, Bladt, and Müller (2022), although in the context of covariate information.

4.4 EM algorithm for the aggregate Markov model

In this section, we give the main contributions of the paper, namely maximum likelihood estimation of micro intensities using the expectation-maximization (EM) algorithm, where we give special attention to the case where the reset property is satisfied along with piecewise constant transition rates. The results of this section naturally extends those of Ahmad, Bladt, and Bladt (2022, Section 3.3).

4.4.1 General EM algorithm

The macro data we observe are N i.i.d. realizations of the macrostate process Z on the generic time interval $[0, T]$. It shall turn out to be useful to represent the data via the associated marked point process $(T_i, Y_i)_{i\in\mathbb{N}_0}$ to keep track of jump times and transitions. This is also the approach of, e.g., Asmussen, Nerman, and Olsson (1996).

Denote with $\mathcal{S}^{(n)} = \big(T^{(n)}, Y^{(n)}\big) = (T_i^{(n)}, Y_i^{(n)})_{i\leq M^{(n)}}$ the n'th observation, $n = 1, \dots, N$, of jump times and transitions of the macrostate process Z, where $M^{(n)}$ is the number of transitions observed, so that $T = \max_{n=1,\dots,N} T_{M^{(n)}}^{(n)}$. Represent all observed data by the vector

$$\boldsymbol{\mathcal{S}} = \Big(\mathcal{S}^{(1)}, \dots, \mathcal{S}^{(N)}\Big).$$

Let $\mathbb{E}_{(\boldsymbol{\pi}_1,\boldsymbol{\theta})}$ denote the expectation under which the Markov jump process $\boldsymbol{X}$ of macro-micro states has transition intensity matrix function $\boldsymbol{M}(\cdot;\boldsymbol{\theta})$ and initial distribution $(\boldsymbol{\pi}_1(0), \boldsymbol{0})$. The EM algorithm for estimation of the micro-level parameter $(\boldsymbol{\pi}_1, \boldsymbol{\theta})$ then consists of initializing with some value $(\boldsymbol{\pi}_1^{(0)}, \boldsymbol{\theta}^{(0)}) \in [0,1]^{d_1} \times \boldsymbol{\Theta}$, and then iteratively compute the conditional expected log-likelihood given macro data under some current parameter $(\boldsymbol{\pi}_1^{(m)}, \boldsymbol{\theta}^{(m)})$, the so-called E-step,

$$(\boldsymbol{\pi}_1, \boldsymbol{\theta}) \mapsto \bar{L}^{(m)}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) = \mathbb{E}_{(\boldsymbol{\pi}_1^{(m)}, \boldsymbol{\theta}^{(m)})}\big[L^{\boldsymbol{X}}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) \,\big|\, \boldsymbol{\mathcal{S}}\big], \quad m \in \mathbb{N}_0, \tag{4.4.1}$$

and then maximize this to update the parameter to $(\boldsymbol{\pi}_1^{(m+1)}, \boldsymbol{\theta}^{(m+1)})$, the so-called M-step. For notational convenience, we write, under some parameter $(\boldsymbol{\pi}_1^{(m)}, \boldsymbol{\theta}^{(m)})$,

$$\bar{\boldsymbol{P}}_i^{(m)}(s,t) = \prod_s^t \Big(\boldsymbol{I} + \boldsymbol{M}_{ii}\big(x; \boldsymbol{\theta}^{(m)}\big)\mathrm{d}x\Big), \quad i \in \mathcal{J}, \tag{4.4.2}$$

for the transition (sub-)probability matrix functions within macrostates, and

$$\boldsymbol{\alpha}^{(m)}(\mathfrak{s}_n) = \boldsymbol{\pi}_1^{(m)}(0) \prod_{\ell=0}^{n-1} \bar{\boldsymbol{P}}_{y_\ell}^{(m)}(t_\ell, t_{\ell+1}) \boldsymbol{M}_{y_\ell y_{\ell+1}}\big(t_{\ell+1}; \boldsymbol{\theta}^{(m)}\big), \tag{4.4.3}$$

for the corresponding defective distribution at time t_n. Also, we denote with $\mathbf{1}_n'$ the row vector of ones with the same dimension as $\boldsymbol{\alpha}^{(m)}(\mathcal{S}^{(n)})$.

To obtain the conditional expected log-likelihood given macro data, we need some conditional expected statistics. For $r \in \{1, \dots, d_1\}$, and $\boldsymbol{i}, \boldsymbol{j} \in E$, $\boldsymbol{j} \neq \boldsymbol{i}$, define

$$\bar{B}_{(1,r)}^{(m)}(0) = \mathbb{E}_{(\boldsymbol{\pi}_1^{(m)},\, \boldsymbol{\theta}^{(m)})}\big[B_{(1,r)}(0) \,\big|\, \boldsymbol{\mathcal{S}}\big], \tag{4.4.4}$$

$$\bar{I}_{\boldsymbol{i}}^{(m)}(s) = \mathbb{E}_{(\boldsymbol{\pi}_1^{(m)},\, \boldsymbol{\theta}^{(m)})}[I_{\boldsymbol{i}}(s) \,|\, \boldsymbol{\mathcal{S}}], \tag{4.4.5}$$

$$\bar{N}_{\boldsymbol{ij}}^{(m)}(s) = \mathbb{E}_{(\boldsymbol{\pi}_1^{(m)},\, \boldsymbol{\theta}^{(m)})}\big[N_{\boldsymbol{ij}}(s) \,\big|\, \boldsymbol{\mathcal{S}}\big]. \tag{4.4.6}$$

Introduce the $d_i \times d_i$ matrix function $\boldsymbol{c}_i^{(m)}$ and the $d_j \times d_i$ matrix function $\boldsymbol{a}_{ij}^{(m,\ell)}$, $i, j \in \mathcal{J}$, $j \neq i$, and $\ell \in \{1, \dots, n\}$, given by

$$\begin{aligned}\boldsymbol{c}_i^{(m)}\big(u; \mathfrak{s}_n\big) &= \sum_{\ell=1}^{n} \mathbb{1}_{[t_{\ell-1},\, t_\ell)}(u)\mathbb{1}_{(y_{\ell-1} = i)} \times \\ &\qquad \bar{\boldsymbol{P}}_i^{(m)}(u, t_\ell)\boldsymbol{\alpha}_\ell^{(m)}(\mathfrak{s}_n)\boldsymbol{\alpha}^{(m)}(\mathfrak{s}_{\ell-1})\bar{\boldsymbol{P}}_i^{(m)}(t_{\ell-1}, u),\\ \boldsymbol{a}_{ij}^{(m,\ell)}\big(u; \mathfrak{s}_n\big) &= \mathbb{1}_{(t_{\ell-1},\, t_\ell]}(u)\mathbb{1}_{(y_{\ell-1} = i,\, y_\ell = j)} \times \\ &\qquad \bar{\boldsymbol{P}}_j^{(m)}(u, t_{\ell+1})\boldsymbol{\alpha}_{\ell+1}^{(m)}(\mathfrak{s}_n)\boldsymbol{\alpha}^{(m)}(\mathfrak{s}_{\ell-1})\bar{\boldsymbol{P}}_i^{(m)}(t_{\ell-1}, u),\end{aligned}$$

where the $d_{y_{\ell-1}}$-dimensional row vector

$$\boldsymbol{\alpha}_\ell^{(m)}(\mathfrak{s}_n) = \boldsymbol{M}_{y_{\ell-1}y_\ell}\big(t_\ell; \boldsymbol{\theta}^{(m)}\big)\left(\prod_{r=\ell}^{n-1} \bar{\boldsymbol{P}}_{y_r}^{(m)}(t_r, t_{r+1})\boldsymbol{M}_{y_r y_{r+1}}\big(t_{r+1}; \boldsymbol{\theta}^{(m)}\big)\right)\mathbf{1}_{d_{y_n}}$$

takes care of sample path probabilities from the ℓ'th jump and onwards. We then have the following main result.

Theorem 4.4.1. *The conditional expected log-likelihood given the macro data $\boldsymbol{\mathcal{S}}$ under the parameter $(\boldsymbol{\pi}_1^{(m)}, \boldsymbol{\theta}^{(m)})$, $m \in \mathbb{N}_0$, is given by*

$$\begin{aligned}\bar{L}^{(m)}(\boldsymbol{\pi}_1, \boldsymbol{\theta}) &= \bar{L}_0^{(m)}(\boldsymbol{\pi}_1) + \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in E\\ \boldsymbol{j} \neq \boldsymbol{i}}} \bar{L}_{\boldsymbol{ij}}^{(m)}(\boldsymbol{\theta}),\\ \bar{L}_0^{(m)}(\boldsymbol{\pi}_1) &= \sum_{r=1}^{d_1} \bar{B}_{(1,r)}^{(m)}(0)\log(\pi_{(1,r)}(0)), \qquad (4.4.7)\\ \bar{L}_{\boldsymbol{ij}}^{(m)}(\boldsymbol{\theta}) &= \int_{(0,T]} \log\big(\mu_{\boldsymbol{ij}}(u; \boldsymbol{\theta})\big)\mathrm{d}\bar{N}_{\boldsymbol{ij}}^{(m)}(u) - \int_0^T \bar{I}_{\boldsymbol{i}}^{(m)}(u)\mu_{\boldsymbol{ij}}(u; \boldsymbol{\theta})\,\mathrm{d}u,\end{aligned}$$

where, for $r \in \{1, \dots, d_1\}$,

$$\bar{B}_{(1,r)}^{(m)}(0) = \sum_{n=1}^{N} \frac{\pi_{(1,r)}^{(m)}(0)\, \boldsymbol{e}_r'\bar{\boldsymbol{P}}_1^{(m)}\big(0, T_1^{(n)}\big)\boldsymbol{\alpha}_1^{(m)}\big(\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n},$$

while for $\boldsymbol{\imath} \in E$ *and* $\check{i} \in \{1, \ldots, d_i\}$, $\check{i} \neq \tilde{i}$,

$$\bar{I}^{(m)}_{\boldsymbol{\imath}}(u) = \sum_{n=1}^{N} \frac{\boldsymbol{e}'_{\check{i}} \boldsymbol{c}^{(m)}_{i}\big(u; \mathcal{S}^{(n)}\big)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n},$$

$$\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{\imath}(i,\check{i})}(u) = \sum_{n=1}^{N} \mu_{\boldsymbol{\imath}(i,\check{i})}\big(u; \boldsymbol{\theta}^{(m)}\big) \frac{\boldsymbol{e}'_{\check{i}} \boldsymbol{c}^{(m)}_{i}\big(u; \mathcal{S}^{(n)}\big)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n}\, \mathrm{d}u,$$

and for $\boldsymbol{\jmath} \in E$, $j \neq i$,

$$\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{\imath}\boldsymbol{\jmath}}(u) = \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mu_{\boldsymbol{\imath}\boldsymbol{\jmath}}\big(u; \boldsymbol{\theta}^{(m)}\big) \frac{\boldsymbol{e}'_{\tilde{j}} \boldsymbol{a}^{(m,\ell)}_{ij}\big(u; \mathcal{S}^{(n)}\big)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n}\, \mathrm{d}\varepsilon_{T^{(n)}_{\ell}}(u),$$

where $\varepsilon_{T^{(n)}_{\ell}}$ *is the Dirac measure in* $T^{(n)}_{\ell}$.

Proof. See Appendix 4.A □

The result shows that in order to develop an EM algorithm for the general aggregate Markov model, one must significantly extend the EM algorithm for general IPHs introduced in Ahmad, Bladt, and Bladt (2022, Appendix A) to computationally more demanding algorithms, even though the conditional sojourn time distributions (4.2.4) follow IPH distributions. This is due to the fact that the general aggregate Markov model admits a path dependency, which is seen from the initial distributions $\boldsymbol{\alpha}$ of the IPH distributions of the conditional sojourn times. We, therefore, require an extension of the already complex algorithm of Ahmad, Bladt, and Bladt (2022, Appendix A) to an algorithm where we keep track of past and future macro paths in the calculation of the conditional expected statistics.

It may be noted from Subsection 4.3.2, cf. also (4.2.7)-(4.2.8), that the above-mentioned computational complexities can be remedied by assuming that the reset property (4.2.6) is satisfied. We, therefore, assume that this is the case for the remainder of the paper to obtain our main algorithms. The general EM algorithm, which we do not implement here, is presented in Appendix 4.B for completeness, as it may have its own merit in future work.

4.4.2 EM algorithm within the reset property

We now assume that the reset property on the form (4.3.6) is satisfied, such that the complete data log-likelihood takes the form (4.3.10). Since we in this case parametrize transition rates in $(\boldsymbol{\theta}, \boldsymbol{\eta}) \in \boldsymbol{\Theta} \times \boldsymbol{H}$, the conditional expected log-likelihood given the macro data $\boldsymbol{\mathcal{S}}$, under some current parameter $(\boldsymbol{\theta}^{(m)}, \boldsymbol{\eta}^{(m)})$ is defined as the map

$$(\boldsymbol{\theta}, \boldsymbol{\eta}) \mapsto \bar{L}^{(m)}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \mathbb{E}_{(\boldsymbol{\theta}^{(m)}, \boldsymbol{\eta}^{(m)})}\big[L^{\boldsymbol{X}}(\boldsymbol{\theta}, \boldsymbol{\eta}) \,\big|\, \boldsymbol{\mathcal{S}}\big], \quad m \in \mathbb{N}_0, \tag{4.4.8}$$

where $\mathbb{E}_{(\boldsymbol{\theta},\boldsymbol{\eta})}$ denotes the expectation under which the Markov jump process $\boldsymbol{X}$ of macro-micro states has transition intensity matrix function $\boldsymbol{M}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta})$ and initial distribution $(\boldsymbol{\pi}_1(0;\boldsymbol{\eta}),\boldsymbol{0})$.

The nature of the reset property allows us to consider each observed macro sojourn independently, and we shall therefore group data into the different macrostates so that computations can be carried out locally within macrostates without the need to take care of past and future macro paths. This is made precise as follows. For $i \in \mathcal{J}$, let

$$M_i = \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{(Y_{\ell-1}^{(n)}=i)}$$

denote the number of sojourns in macrostate i observed, and furthermore let $\boldsymbol{\mathcal{T}}_i = \big(\mathcal{T}_i^{(1)},\ldots,\mathcal{T}_i^{(M_i)}\big)$ be the set of macrostate i observations given by

$$\begin{aligned}\boldsymbol{\mathcal{T}}_i &= \Big\{ \Big(T_{\ell-1}^{(n)},\, Y_{\ell-1}^{(n)},\, T_\ell^{(n)},\, Y_\ell^{(n)}\Big) \,\Big|\, n=1,\ldots,N,\, \ell=1,\ldots,M^{(n)} \text{ s.t } Y_{\ell-1}^{(n)}=i\Big\} \\ &= \Big\{ \Big(R_i^{(n)},\, i,\, \tau_i^{(n)},\, Z_i^{(n)}\Big)\Big\}_{n\in\{1,\ldots,M_i\}}. \end{aligned} \tag{4.4.9}$$

Then $\boldsymbol{\mathcal{T}}_i$ contains data points for macrostate i, consisting of time of entries $R_i^{(n)}$ into the macrostate, jump times $\tau_i^{(n)}$ out of the state, and macrostates $Z_i^{(n)}$ jumped to at time $\tau_i^{(n)}$. Similar type of data representations are made in Breuer (2002).

For a generic realization $\sqcup_i = (r_i, i, \tau_i, z_i)$ of $\mathcal{T}_i^{(n)}$, the matrix function $\boldsymbol{c}_i^{(m)}$ and defective distribution $\boldsymbol{\alpha}^{(m)}$ now satisfy, for $u \in (r_i, \tau_i]$,

$$\begin{aligned}\boldsymbol{c}_i^{(m)}(u;\sqcup_i) &= \bar{\boldsymbol{P}}_i^{(m)}(u,\tau_i)\boldsymbol{\beta}_{iz_i}\big(\tau_i;\boldsymbol{\theta}^{(m)}\big)\boldsymbol{\pi}_i\big(r_i;\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}(r_i,u), \\ \boldsymbol{\alpha}^{(m)}(\sqcup_i)\boldsymbol{1}_{d_{z_i}} &= \boldsymbol{\pi}_i\big(r_i;\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}(r_i,\tau_i)\boldsymbol{\beta}_{iz_i}\big(\tau_i;\boldsymbol{\theta}^{(m)}\big).\end{aligned} \tag{4.4.10}$$

Concerning jumps between macrostates, introduce the aggregated conditional expected statistics

$$\bar{N}_{\boldsymbol{i}j}(s) = \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}[N_{\boldsymbol{i}j}(s)\,|\,\boldsymbol{\mathcal{S}}] \qquad \text{and} \qquad \bar{N}_{\boldsymbol{i}}(s) = \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}[N_{\boldsymbol{i}}(s)\,|\,\boldsymbol{\mathcal{S}}],$$

where the aggregated statistics $N_{\boldsymbol{i}j}$ and $N_{\boldsymbol{i}}$ are given in (4.3.9). Furthermore, the conditional expected statistics within macrostates, $\bar{I}_{\boldsymbol{i}}^{(m)}$ and $\bar{N}_{\boldsymbol{i}(i,\check{\imath})}^{(m)}$, are given as in the general case in (4.4.5)-(4.4.6), but where the expectation is taken under $(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})$, i.e. with the operator $\mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}$. We now have the following result.

Corollary 4.4.2. *Suppose that the reset property* (4.2.6) *holds. Then the conditional expected log-likelihood* (4.4.8) *is given by*

$$\bar{L}^{(m)}(\boldsymbol{\theta},\boldsymbol{\eta}) = \sum_{\boldsymbol{i}\in E}\left(\bar{L}_{\boldsymbol{i}}^{(m)}(\boldsymbol{\eta}) + \sum_{\substack{\check{\imath}=1\\ \check{\imath}\neq i}}^{d_i}\bar{L}_{\boldsymbol{i}(i,\check{\imath})}^{(m)}(\boldsymbol{\theta}) + \sum_{\substack{j\in\mathcal{J}\\ j\neq i}}\bar{L}_{\boldsymbol{i}j}^{(m)}(\boldsymbol{\theta})\right), \tag{4.4.11}$$

where $\bar{L}^{(m)}_{\boldsymbol{i}(i,\check{i})}$ *is as in* (4.4.7)*, while*

$$\begin{aligned}\bar{L}^{(m)}_{\boldsymbol{i}}(\boldsymbol{\eta}) &= \int_{[0,T]} \log\left(\pi_{\boldsymbol{i}}(u;\boldsymbol{\eta})\right)\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}}(u),\\ \bar{L}^{(m)}_{\boldsymbol{i}j}(\boldsymbol{\theta}) &= \int_{(0,T]} \log\left(\beta_{\boldsymbol{i}j}(u;\boldsymbol{\theta})\right)\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}j}(u) - \int_0^T \bar{I}^{(m)}_{\boldsymbol{i}}(u)\beta_{\boldsymbol{i}j}(u;\boldsymbol{\theta})\,\mathrm{d}u,\end{aligned} \tag{4.4.12}$$

but where the conditional expected statistics are given by, for $\boldsymbol{i} \in E$ *and* $\check{i} \in \{1,\dots,d_i\}$, $\check{i} \neq \tilde{i}$,

$$\begin{aligned}\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}}(u) &= \sum_{n=1}^{M_i} \frac{\pi_{\boldsymbol{i}}\left(u;\boldsymbol{\eta}^{(m)}\right)\boldsymbol{e}'_{\tilde{i}}\bar{\boldsymbol{P}}^{(m)}_i\left(u,\tau^{(n)}_i\right)\boldsymbol{\beta}_{iZ^{(n)}_i}\left(\tau^{(n)}_i;\boldsymbol{\theta}^{(m)}\right)}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n}\,\mathrm{d}\varepsilon_{R^{(n)}_i}(u),\\ \bar{I}^{(m)}_{\boldsymbol{i}}(u) &= \sum_{n=1}^{M_i} \frac{\boldsymbol{e}'_{\tilde{i}}\boldsymbol{c}^{(m)}_i\left(u;\mathcal{T}^{(n)}_i\right)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n},\\ \mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}(i,\check{i})}(u) &= \sum_{n=1}^{M_i} \mu_{\boldsymbol{i}(i,\check{i})}\left(u;\boldsymbol{\theta}^{(m)}\right)\frac{\boldsymbol{e}'_{\check{i}}\boldsymbol{c}^{(m)}_i\left(u;\mathcal{T}^{(n)}_i\right)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n}\,\mathrm{d}u,\end{aligned} \tag{4.4.13}$$

while for $j \in \mathcal{J}$, $j \neq i$,

$$\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}j}(u) = \sum_{n=1}^{M_i} \mathbb{1}_{(Z^{(n)}_i=j)}\frac{\boldsymbol{\pi}_i\left(R^{(n)}_i;\boldsymbol{\eta}^{(m)}\right)\bar{\boldsymbol{P}}^{(m)}_i\left(R^{(n)}_i,u\right)\boldsymbol{e}_{\tilde{i}}\beta_{\boldsymbol{i}j}\left(u;\boldsymbol{\theta}^{(m)}\right)}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n}\,\mathrm{d}\varepsilon_{\tau^{(n)}_i}(u). \tag{4.4.14}$$

Proof. See Appendix 4.A. □

Remark 4.4.3. The conditional expected log-likelihood (4.4.11) can be seen to have close relations to the conditional expected log-likelihood of Ahmad, Bladt, and Bladt (2022, Theorem 3.4). Indeed, consider e.g.

$$\int_0^T \bar{I}^{(m)}_{\boldsymbol{i}}(u)\mu_{\boldsymbol{i}\boldsymbol{j}}(u;\boldsymbol{\theta})\mathrm{d}u = \sum_{n=1}^{M_i}\int_{R^{(n)}_i}^{\tau^{(n)}_i} \frac{\boldsymbol{e}'_{\tilde{i}}\boldsymbol{c}^{(m)}_i\left(u;\mathcal{T}^{(n)}_i\right)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n}\mu_{\boldsymbol{i}\boldsymbol{j}}(u;\boldsymbol{\theta})\,\mathrm{d}u.$$

Looking at a single term on the right-hand side and applying the substitution $v = u - R^{(n)}_i$ to the integral, we see that it equals

$$\int_0^{\tau^{(n)}_i - R^{(n)}_i} \frac{\boldsymbol{e}'_{\tilde{i}}\boldsymbol{c}^{(m)}_i\left(v + R^{(n)}_i;\mathcal{T}^{(n)}_i\right)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\left(\mathcal{T}^{(n)}_i\right)\mathbf{1}_n}\mu_{\boldsymbol{i}\boldsymbol{j}}(v + R^{(n)}_i;\boldsymbol{\theta})\,\mathrm{d}v,$$

where the shifted versions of $\boldsymbol{c}^{(m)}_i$ equals

$$\prod_{v}^{\tau_i^{(n)}-R_i^{(n)}} \Big(\boldsymbol{I}+\boldsymbol{M}_{ii}\big(x+R_i^{(n)};\boldsymbol{\theta}^{(m)}\big)\mathrm{d}x\Big)\boldsymbol{\beta}_{iZ_i^{(n)}}\big(\tau_i^{(n)};\boldsymbol{\theta}^{(m)}\big)\times$$
$$\boldsymbol{\pi}_i\big(R_i^{(n)};\boldsymbol{\eta}^{(m)}\big)\prod_{0}^{v}\Big(\boldsymbol{I}+\boldsymbol{M}_{ii}\big(x+R_i^{(n)};\boldsymbol{\theta}^{(m)}\big)\mathrm{d}x\Big).$$

Performing similar type of manipulations for the other terms in (4.4.13)-(4.4.14), we see that each term in the conditional expected log-likelihood (4.4.11), corresponding to each macro sojourn $n \in \{1,\dots,M_i\}$, $i \in \mathcal{J}$, equals the conditional expected log-likelihood of Ahmad, Bladt, and Bladt (2022, Theorem 3.4), if in the latter we have

- A single IPH observation $\tau_i^{(n)} - R_i^{(n)}$
- Initial distribution $\boldsymbol{\pi}_i\big(R_i^{(n)};\boldsymbol{\eta}^{(m)}\big)$
- Sub-intensity matrix function $x \mapsto \boldsymbol{M}_{ii}\big(x+R_i^{(n)};\boldsymbol{\theta}^{(m)}\big)$
- Exit rate vector function $x \mapsto \boldsymbol{\beta}_{iZ_i^{(n)}}\big(x+R_i^{(n)};\boldsymbol{\theta}^{(m)}\big)$

on the state-space $\{1,\dots,d_i\}$ of transient states, and with parameter space $\boldsymbol{\Theta}$. $\triangle$

It follows from Corollary 4.4.2 and Remark 4.4.3 that the E-step of the EM algorithm for the aggregate Markov model with the reset property can be formulated in terms of, and executed by, the E-step of the EM algorithm in Ahmad, Bladt, and Bladt (2022, Appendix A). The computational demand of performing the estimation procedure, in this case, is therefore comparable to those for general IPHs. As explained in Ahmad, Bladt, and Bladt (2022, Subsection 3.3), these computational demands are generally much higher than those of, e.g., Albrecher, Bladt, and Yslas (2022) and Asmussen, Nerman, and Olsson (1996), and assuming piecewise constant transition rates may therefore be of significant advantage. We shall follow this approach for the remainder of the paper to obtain our main algorithm, using the setup made in Subsection 4.3.4. As for the general case, we present the EM algorithm resulting from this section in Appendix 4.B for completeness, as it may have its own merit in future work.

4.4.3 EM algorithm with piecewise constant transition rates within the reset property

We now consider the simplifications arising from assuming that the transition intensity matrix function $\boldsymbol{M}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta})$ is piecewise constant on the form (4.2.9) along with the reset property (4.3.18) being satisfied. Since the resulting complete data log-likelihood, (4.3.10) with (4.3.14) and (4.3.21)-(4.3.22), is linear in the sufficient statistics, we see that for the E-step, it now suffices to compute the conditional expected sufficient statistics,

$$\begin{aligned}
\bar{B}_{\boldsymbol{i}}^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}[B_{\boldsymbol{i}}(k)\,|\,\boldsymbol{\mathcal{S}}],\\
\bar{E}_{\boldsymbol{i}}^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}[E_{\boldsymbol{i}}(k)\,|\,\boldsymbol{\mathcal{S}}],\\
\bar{O}_{\boldsymbol{i}(i,\check{\imath})}^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}\Big[O_{\boldsymbol{i}(i,\check{\imath})}(k)\,\Big|\,\boldsymbol{\mathcal{S}}\Big],\\
\bar{O}_{\boldsymbol{i}j}^{(m)}(k) &= \mathbb{E}_{(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})}[O_{\boldsymbol{i}j}(k)\,|\,\boldsymbol{\mathcal{S}}],
\end{aligned} \tag{4.4.15}$$

and then the M-step regarding the update of $\boldsymbol{\theta}$ simplifies to a Poisson regression, while the update of $\boldsymbol{\eta}$ simplifies to a multinomial regression, as described in Subsection 4.3.4, but where the sufficient statistics are replaced by their conditional expectations computed in the E-step.

The transition (sub-)probability matrices within macrostates (4.4.2) and corresponding defective distribution (4.4.3) under the parameter $(\boldsymbol{\theta}^{(m)},\boldsymbol{\eta}^{(m)})$ are now on the form (cf. (4.2.10)-(4.2.11)):

$$\bar{\boldsymbol{P}}_i^{(m)}(s,t) = \mathrm{e}^{\boldsymbol{M}_{ii}^{k(s)}(\boldsymbol{\theta}^{(m)})\left(s_{k(s)}-s\right)}\left(\prod_{\ell=k(s)+1}^{k(t)-1}\mathrm{e}^{\boldsymbol{M}_{ii}^{\ell}(\boldsymbol{\theta}^{(m)})(s_\ell-s_{\ell-1})}\right)\mathrm{e}^{\boldsymbol{M}_{ii}^{k(t)}(\boldsymbol{\theta}^{(m)})\left(t-s_{k(t)}\right)},$$

$$\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}_i^{(n)}\big)\mathbf{1}_n = \boldsymbol{\pi}_i^{k_i^{(n-)}}\big(\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}\big(R_i^{(n)},\tau_i^{(n)}\big)\boldsymbol{\beta}_{iZ_i^{(n)}}^{k_i^{(n+)}}\big(\boldsymbol{\theta}^{(m)}\big), \tag{4.4.16}$$

where we recall that $k(x)$, for $x \geq 0$, equals the unique $k \in \{1,\dots,K\}$ satisfying that $x \in (s_{k-1}, s_k]$; for notational convenience, we put $k_i^{(n-)} = k\big(R_i^{(n)}\big)$ and $k_i^{(n+)} = k\big(\tau_i^{(n)}\big)$. The conditional expected sufficient statistics (4.4.15) then follow immediately from the more general results of Corollary 4.4.2.

Corollary 4.4.4. *Suppose that the transition intensity matrix function* $\boldsymbol{M}(\cdot;\boldsymbol{\theta},\boldsymbol{\eta})$ *is piecewise constant on the form* (4.2.9)*, and that the reset property* (4.3.18) *is satisfied. Then the conditional expected sufficient statistics* (4.4.15) *are given by,*

$$\bar{B}_{\boldsymbol{i}}^{(m)}(k) = \sum_{n=1}^{M_i}\mathbb{1}_{(k_i^{(n-)}=k)}\frac{\pi_{\boldsymbol{i}}^{k_i^{(n-)}}\big(\boldsymbol{\eta}^{(m)}\big)\,\boldsymbol{e}_{\tilde{\imath}}'\bar{\boldsymbol{P}}_i^{(m)}\big(R_i^{(n)},\tau_i^{(n)}\big)\boldsymbol{\beta}_{iZ_i^{(n)}}^{k_i^{(n+)}}\big(\boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}_i^{(n)}\big)\mathbf{1}_n},$$

$$\bar{E}_{\boldsymbol{i}}^{(m)}(k) = \sum_{n=1}^{M_i}\frac{\displaystyle\int_{\tau_{i|k-1}^{(n)}}^{\tau_{i|k}^{(n)}}\boldsymbol{e}_{\tilde{\imath}}'\boldsymbol{c}_i^{(m)}\big(u;\mathcal{T}_i^{(n)}\big)\boldsymbol{e}_{\tilde{\imath}}\,\mathrm{d}u}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}_i^{(n)}\big)\mathbf{1}_n},$$

$$\bar{O}^{(m)}_{\boldsymbol{i}(i,\check{i})}(k) = \sum_{n=1}^{M_i} \frac{\int_{\tau^{(n)}_{i|k-1}}^{\tau^{(n)}_{i|k}} \boldsymbol{e}'_{\check{i}} \boldsymbol{c}^{(m)}_i\big(u; \mathcal{T}^{(n)}_i\big)\boldsymbol{e}_{\tilde{i}}\,\mathrm{d}u}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}^{(n)}_i\big)\mathbf{1}_n} \mu^k_{\boldsymbol{i}(i,\check{i})}\big(\boldsymbol{\theta}^{(m)}\big),$$

$$\bar{O}^{(m)}_{\boldsymbol{i}j}(k) = \sum_{n=1}^{M_i} \mathbb{1}_{(Z^{(n)}_i = j)} \mathbb{1}_{(k^{(n+)}_i = k)} \frac{\boldsymbol{\pi}^{k^{(n-)}_i}_i\big(\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}^{(m)}_i\big(R^{(n)}_i, \tau^{(n)}_i\big)\boldsymbol{e}_{\tilde{i}}\beta^{k^{(n+)}_i}_{\boldsymbol{i}j}\big(\boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}^{(n)}_i\big)\mathbf{1}_n},$$

with $\bar{\boldsymbol{P}}^{(m)}$ *and* $\boldsymbol{\alpha}^{(m)}$ *as in* (4.4.16)*, and where*

$$\tau^{(n)}_{i|k} = \Big(s_k \vee R^{(n)}_i\Big) \wedge \tau^{(n)}_i.$$

Proof. By inserting the expressions for the sufficient statistics in the complete data case, (4.3.13) combined with (4.3.20), into (4.4.15), the result follows immediately from (4.4.13)-(4.4.14) in Corollary 4.4.2. □

By employing the same techniques as in Remark 4.4.3 on the conditional expected statistics of Corollary 4.4.4, we find that the E-step in this case can be written in terms of the E-step of Ahmad, Bladt, and Bladt (2022, Algorithm 1), with analogue modifications of shifting all inputs from time 0 to the time of entries into the macrostate, $R^{(n)}_i$, and so the computational demand should be comparable to the estimation of IPHs with piecewise constant transition rates. In particular, the sub-intensity matrix function is shifted, and so one must accordingly shift the grid points on which it is piecewise constant. This is a conceptually different modification than that of the more general case of Subsection 4.4.2.

The complete EM algorithm for the aggregate Markov model with piecewise constant transition rates within the reset property is presented in Algorithm 4.1 below. We implement this algorithm and show a numerical example of its applicability in the following section.

Algorithm 4.1 EM algorithm for the aggregate Markov model with the reset property and piecewise constant transition rates

Input: Initial parameters $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\eta}^{(0)}) \in \boldsymbol{\Theta} \times \boldsymbol{H}$, and for each macrostate $i \in \mathcal{J}$, data points within the macrostate,

$$\boldsymbol{\mathcal{T}}_i = \left\{ \left(r_i^{(n)}, \tau_i^{(n)}, z_i^{(n)} \right) \right\}_{n \in \{1, \ldots, m_i\}},$$

consisting of time of entries $r_i^{(n)}$ into the macrostate, jump times $\tau_i^{(n)}$ out of the state, and macrostate $z_i^{(n)}$ jumped to at time $\tau_i^{(n)}$.

0) Set $m := 0$
1) *E-step:* For each macrostate $i \in \mathcal{J}$,
   - For each sojourn $n \in \{1, \ldots, m_i\}$,
     i) Set $k_i^{(n-)} := k\big(r_i^{(n)}\big)$ and $k_i^{(n+)} := k\big(\tau_i^{(n)}\big)$.
     ii) Run the E-step of Ahmad, Bladt, and Bladt (2022, Algorithm 1) with
        – Grid points $0 = \tilde{s}_0 < \tilde{s}_1 < \cdots < \tilde{s}_{K - k_i^{(n-)}} < \tilde{s}_{K - k_i^{(n-)} + 1} = \infty$, where
        $$\tilde{s}_k = s_{k_i^{(n-)} + k - 1} - r_i^{(n)}, \qquad k \in \big\{1, \ldots, K - k_i^{(n-)} + 1\big\}.$$
        – State-space of transient states $\{1, \ldots, d_i\}$
        – A single IPH observation $\tau_i^{(n)} - r_i^{(n)}$
        – Initial distribution $\boldsymbol{\pi}_i^{k_i^{(n-)}}\big(\boldsymbol{\eta}^{(m)}\big)$
        – Sub-intensity matrix function $x \mapsto \boldsymbol{M}_{ii}\big(x + r_i^{(n)}; \boldsymbol{\theta}^{(m)}\big)$
        – Exit rate vector function $x \mapsto \boldsymbol{\beta}_{iz_i^{(n)}}\big(x + r_i^{(n)}; \boldsymbol{\theta}^{(m)}\big)$

        which, across $k \in \big\{1, \ldots, K - k_i^{(n-)} + 1\big\}$, gives the output:
        – Expected statistics for the initial state: $\bar{B}_{\boldsymbol{i}}^{(n,m)}\big(k_i^{(n-)}\big)$
        – Expected exposures: $\bar{E}_{\boldsymbol{i}}^{(n,m)}\big(k_i^{(n-)} + k - 1\big)$
        – Expected occurrences for transient states: $\bar{O}_{\boldsymbol{i}(i,\check{i})}^{(n,m)}\big(k_i^{(n-)} + k - 1\big)$
        – Expected occurrences for the 'absorbing' state: $\bar{O}_{\boldsymbol{i}z_i^{(n)}}^{(n,m)}\big(k_i^{(n+)}\big)$

iii) Compute total expected sufficient statistics, for $k \in \{1, \ldots, K\}$,

$$\bar{B}_{\boldsymbol{i}}^{(m)}(k) = \sum_{n=1}^{M_i} \bar{B}_{\boldsymbol{i}}^{(n,m)}(k), \qquad \bar{O}_{\boldsymbol{i}(i,\check{\imath})}^{(m)}(k) = \sum_{n=1}^{M_i} \bar{O}_{\boldsymbol{i}(i,\check{\imath})}^{(n,m)}(k),$$

$$\bar{E}_{\boldsymbol{i}}^{(m)}(k) = \sum_{n=1}^{M_i} \bar{E}_{\boldsymbol{i}}^{(n,m)}(k), \qquad \bar{O}_{\boldsymbol{i}z_i^{(n)}}^{(m)}(k) = \sum_{n=1}^{M_i} \bar{O}_{\boldsymbol{i}z_i^{(n)}}^{(n,m)}(k).$$

2) *M-step:* Update the parameters:

i) Compute $\hat{\boldsymbol{\eta}}^{(m+1)}$ as the MLE of the regressions

$$\Big(\bar{B}_{(i,1)}^{(m)}(k), \ldots, \bar{B}_{(i,d_i)}^{(m)}(k)\Big) \sim \text{Multinomial}\big(B_i(k), \boldsymbol{\pi}_i^k(\boldsymbol{\eta})\big),$$

across $k \in \{0, 1, \ldots, K\}$ and $i \in \mathcal{J}$.

ii) Compute $\hat{\boldsymbol{\theta}}^{(m+1)}$ as the MLE of the regressions

$$\bar{O}_{\boldsymbol{i}(i,\check{\imath})}^{(m)}(k) \sim \text{Pois}\Big(\mu_{\boldsymbol{i}(i,\check{\imath})}^{k}(\boldsymbol{\theta})\bar{E}_{\boldsymbol{i}}^{(m)}(k)\Big),$$

$$\bar{O}_{\boldsymbol{i}j}^{(m)}(k) \sim \text{Pois}\Big(\beta_{\boldsymbol{i}j}^{k}(\boldsymbol{\theta})\bar{E}_{\boldsymbol{i}}^{(m)}(k)\Big),$$

across $k \in \{1, \ldots, K\}$, $\boldsymbol{i} \in E$, $\check{\imath} \in \{1, \ldots, d_i\}$, $\check{\imath} \neq \tilde{\imath}$, and $j \in \mathcal{J}$.

3) Set $m := m + 1$ and GOTO 1), unless a stopping rule is satisfied.

***Output**: Fitted parameters* $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$.

4.5 Numerical example

In this section, we present a numerical example illustrating the methods developed in Section 4.4. The purpose of the example is to let the EM algorithm fit micro intensities based on macro data, which are simulated from a time-inhomogeneous semi-Markov model already used in the context of multi-state life insurance, see, e.g., Hoem (1972), Helwich (2008), Christiansen (2012), and Buchardt, Møller, and Schmidt (2015) for these type of models. Since the aggregate Markov model with the reset property exactly admits a time-inhomogeneous semi-Markovian structure (see Subsection 4.2 in Ahmad, Bladt, and Furrer, 2022), we are able to apply algorithms within this special case to fit an aggregate Markov model with the reset property that sufficiently captures the duration effects appearing in these kinds of models.

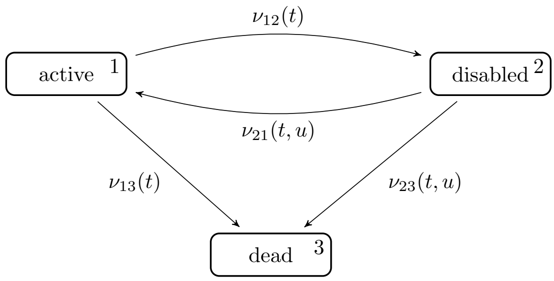

Figure 4.1: *The time-inhomogeneous semi-Markovian disability model with recoveries where the macro data is simulated from.*

We simulate 10,000 paths of macro data from the three-state disability model depicted in Figure 4.1, where we consider a 30-year old male being active upon initiation. For the time and duration-dependent transition rates ν_{ij}, we use a set of transition rates that are based on rates employed by a large Danish life insurance company for males, which has been reported to and published by the Danish Financial Supervisory Authority. They are given by, for ages $t \in [30, 110]$ and durations $u \leq t$,

$\nu_{13}(\cdot)$: The 2012 edition of the Danish FSA's longevity benchmark,

$$\nu_{12}(t) = \begin{cases} e^{72.539-10.669t+0.534t^2-0.0128t^3+1.4922\cdot 10^{-4}t^4-6.8007\cdot 10^{-7}t^5} & \text{for } t \leq 67 \\ 0.0009687435 & \text{for } t > 67 \end{cases}$$

$$\nu_{21}(t,u) = \begin{cases} e^{-0.9148875-0.0309126t+4.8715347u} & \text{for } u \leq 0.2291667 \\ e^{0.3766531-0.0309126t-0.7642786u} & \text{for } u \in (0.2291667, 2] \\ e^{-0.4808001-0.0309126t-0.335552u} & \text{for } u \in (2, 5] \\ e^{-0.042168-0.092455t} & \text{for } u > 5 \end{cases} \tag{4.5.1}$$

$$\nu_{23}(t,u) = \begin{cases} e^{-6.1057464+0.0635736t-0.2891195u} & \text{for } u \leq 5 \\ e^{-11.9169277+0.1356766t} & \text{for } u > 5 \end{cases}$$

Since this model only contains duration dependence regarding transitions from the disabled state, we fit the aggregate Markov model depicted in Figure 4.2 to the simulated data, where only the disabled state contains microstates, so that $d_1 = d_3 = 1$ and $d_2 \geq 1$. This model always has the reset property satisfied, and it allows us to capture the duration effects regarding transitions from the disabled state.

Algorithm 4.1 is then used to fit piecewise constant transition rates and initial distributions on the micro level in the disabled state for a varying number of disability microstates, $d_2 \in \{1, 2, 3, 5, 7, 10\}$. We use piecewise constant transition

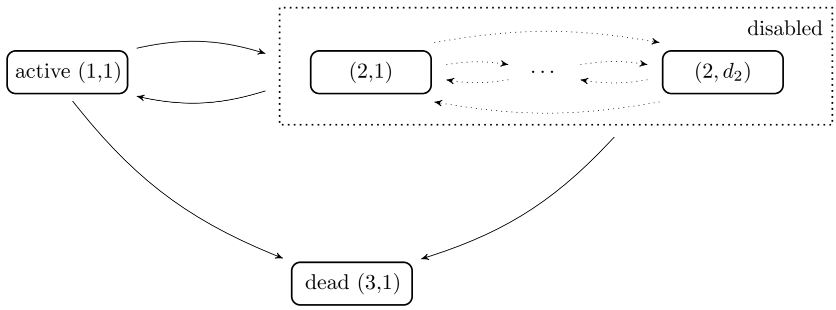

Figure 4.2: *Disability model with d_2 unobservable disability microstates.*

rates on the grid $\{30, 36, 37, \dots, 88, 110\}$, and the rates between microstates as well as exit rates are parametrized similarly as the true rates (4.5.1), that is, (4.3.17) with $q = 1$ and $g^{(1)}(s) = s$, such that Poisson GLMs of occurrences against time (with logarithmic link function and log-exposure as offset) is carried out in the M-step; this implies that logarithmic transition rates are (piecewise) linear in time. For the initial distribution, we use (4.3.23) so that multinomial logistic regressions of the number of initiations against time are carried out in the M-step. The resulting fits of the transition rates in the disabled state on the micro level, as well as initial distributions and exit rates, can be found in Figure 4.6 for the case $d_2 = 10$.

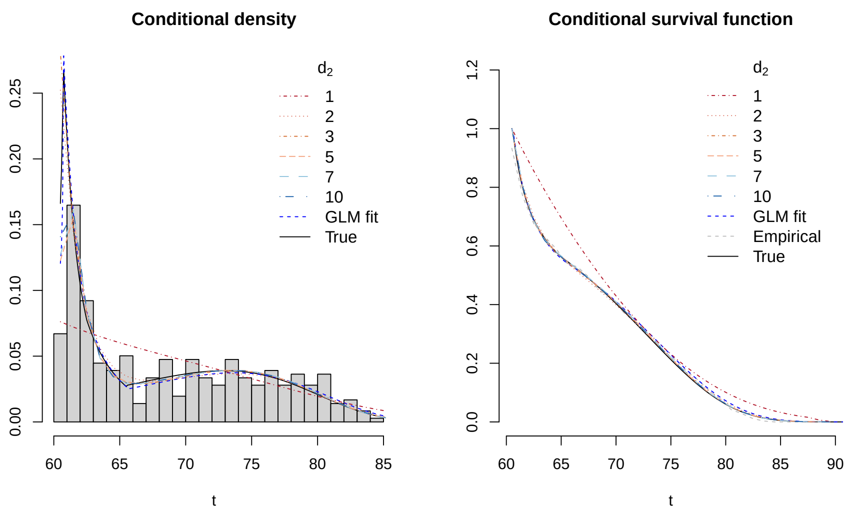

Figure 4.3: *Conditional densities (left) and survival functions (right) of the sojourn time in the disabled state given entry at time $s = 60.5$, for different number of disability microstates, d_2, along with the GLM fit as well as true and empirical distributions. The case $d_2 = 1$ corresponds to a Markov chain.*

We then compare our model fits with the true model, of course, but to avoid statistical noise and also allow for comparison to more classic methods, we also carry out a semi-Markovian GLM fit on the macro data, that is, a Poisson GLM fit of occurrences against age and duration with log-link function and log-exposure as offset. Here we use segmented regressions in the duration dimension to obtain the parametrization (4.5.1).

In Figure 4.3 we show for $s = 60.5$ the resulting fits of the conditional survival function and density of the sojourn time distribution in the disabled state given entry into the state at time s, respectively given by

$$t \mapsto \boldsymbol{\pi}_2(s;\hat{\boldsymbol{\eta}})\bar{\boldsymbol{P}}_2(s,t;\hat{\boldsymbol{\theta}})\mathbf{1}_{d_2} \qquad \text{and} \qquad t \mapsto \boldsymbol{\pi}_2(s;\hat{\boldsymbol{\eta}})\bar{\boldsymbol{P}}_2(s,t;\hat{\boldsymbol{\theta}})\Big(\boldsymbol{\beta}_{21}(t;\hat{\boldsymbol{\theta}}) + \boldsymbol{\beta}_{23}(t;\hat{\boldsymbol{\theta}})\Big)$$

against their empirical counterparts as well as (cf., e.g., Hoem, 1972; Helwich, 2008; Christiansen, 2012):

$$t \mapsto \mathrm{e}^{-\int_s^t \nu_{2\cdot}(v,v-s)\,\mathrm{d}v} \qquad \text{and} \qquad t \mapsto \mathrm{e}^{-\int_s^t \nu_{2\cdot}(v,v-s)\,\mathrm{d}v}(\nu_{21}(t,t-s) + \nu_{23}(t,t-s))$$

for the true model and the GLM fits. We see that a single microstate, corresponding to a Markov chain, does not fit very well to the data, which is to be expected as this model is not able to capture the duration effects appearing in the true model. However, we see that by adding microstates to the disabled state, we are able to fit the distributions with a high accuracy already with 2-3 microstates. A challenging part of the density close to the origin (corresponding to small durations) as well as around discontinuity points requires many more microstates compared to the GLM fit, though, but this does not seem to carry over to the corresponding survival function, where our model fits performs very well.

While these survival functions and densities focus on sojourn times and, therefore, not on specific transitions out of the state, we may further examine how the actual transition rates are fitted; these play an important role in the context of multi-state life insurance. Indeed, according to Ahmad, Bladt, and Furrer (2022, Subsection 4.2), we have fitted an aggregate Markov model with the reset property that admits a time-inhomogeneous semi-Markovian structure with the following time and duration-dependent transition rates out of the disabled state:

$$(t,u) \mapsto \frac{\boldsymbol{\pi}_2(t-u;\hat{\boldsymbol{\eta}})\bar{\boldsymbol{P}}_2(t-u,t;\hat{\boldsymbol{\theta}})}{\boldsymbol{\pi}_2(t-u;\hat{\boldsymbol{\eta}})\bar{\boldsymbol{P}}_2(t-u,t;\hat{\boldsymbol{\theta}})\mathbf{1}_{d_2}}\boldsymbol{\beta}_{2j}(t;\hat{\boldsymbol{\theta}}), \qquad j \in \mathcal{J},\ j \neq 2. \tag{4.5.2}$$

In Figure 4.4, we examine this as a function of time, t, for a fixed duration $u = 1$, and compare it with the true rates $t \mapsto \nu_{2j}(t,1)$, as well as its empirical counterpart, namely empirical occurrence-exposure rates on the macro level in the disabled state. Here, we also see how badly the Markov chain case of a single microstate fits, but more importantly, we see that it requires more microstates to be able to fit the actual transition rates more accurately than we saw with the sojourn time distributions. However, since we see a higher deviation away from the true model for the GLM fit, especially for recoveries, this might also be due to statistical noise. Indeed, except for a very small amount of data for recoveries at young ages, we see fits with a high number of microstates that performs at least as well as the GLM fit.

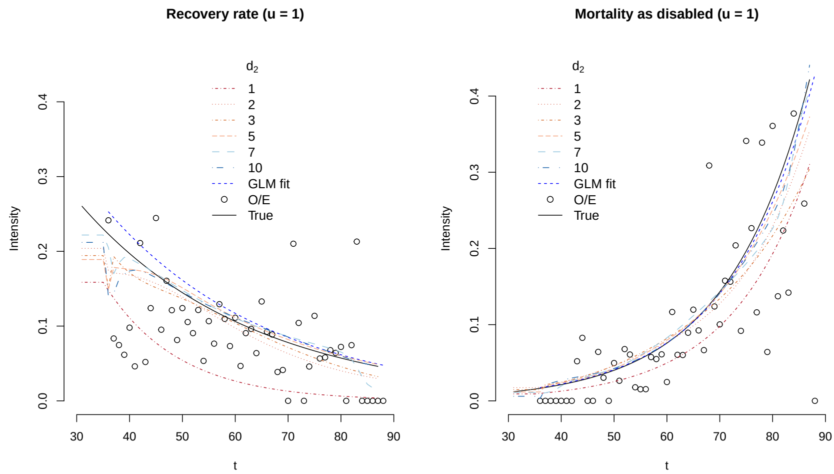

Figure 4.4: *Estimated transition rates* (4.5.2) *for recoveries (left) and deaths as disabled (right) as a function of time, t, with a fixed duration $u = 1$ for different number of disability microstates, d_2, along with the GLM fit as well as the true rates, ν_{2j}, and the empirical occurrence-exposure rates.*

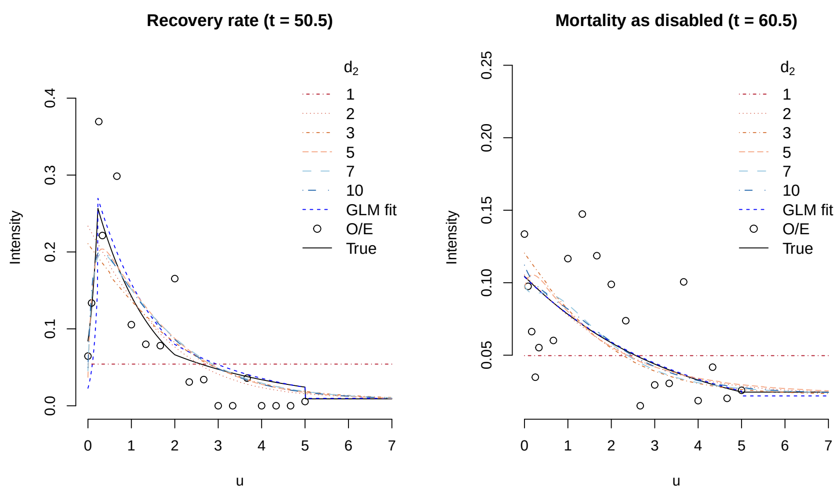

Figure 4.5: *Estimated transition rates* (4.5.2) *for recoveries (left) and deaths as disabled (right) as a function of duration, u, with a fixed time, t, for different number of disability microstates, d_2, along with the GLM fit as well as the true rates, ν_{2j}, and the empirical occurrence-exposure rates.*

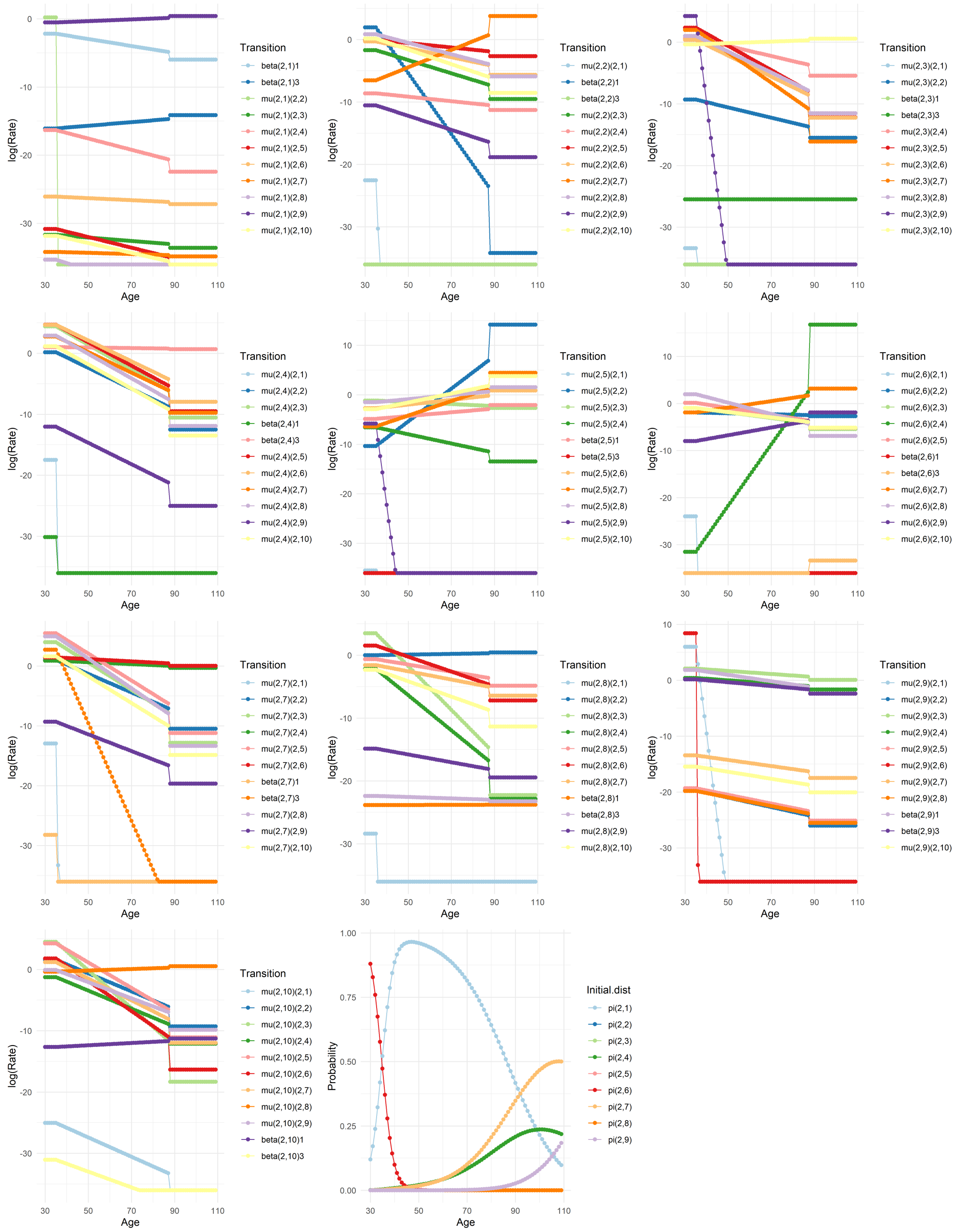

Figure 4.6: *Fitted transition rates on micro level in the disabled state, $\mu_{(2,\tilde{\imath})(2,\check{\imath})}(\cdot;\hat{\boldsymbol{\theta}})$, as well as the initial distribution $\pi_{(2,\tilde{\imath})}(\cdot;\hat{\boldsymbol{\eta}})$ and exit rates $\beta_{(2,\tilde{\imath})j}(\cdot;\hat{\boldsymbol{\theta}})$, for the case $d_2 = 10$.*

To round of the analysis, we consider in Figure 4.4 the fitted rates (4.5.2) as a function of the duration u for a fixed time, t. Due to the aforementioned lack of data, we focus on $t = 50.5$ for the recovery rate and $t = 60.5$ for the mortality as disabled, which are ages with sufficient data. Here, we also see the lower accuracy in the fits close to the origin, as with the conditional densities, which is particularly the case for the recovery rate. The irregularity of the true rates in this region seems to require many microstates for the model to fully capture this kind of duration effect, though the GLM fit also seems to have the same problems, so the performance of the two fits may also be comparable here. For the mortality as disabled, however, we see that the GLM fits much better for durations smaller than five years, while our model then seems to fit better for higher durations. This conforms with the observations made for the conditional densities.

Acknowledgments

We should like to thank Christian Furrer for valuable suggestions and fruitful discussions during this work.

4.A Proofs

In this section, we present the proofs of Theorem 4.4.1 and Corollary 4.4.2. To prove Theorem 4.4.1, we need the following lemma which builds upon Ahmad, Bladt, and Bladt (2022, Lemma 3.2). The result relates to similar conditional distributions considered in Norberg (1991) and Hoem (1969b), but where we include past and future jump times in the conditioning.

Lemma 4.A.1. *Let $\boldsymbol{X} = \{\boldsymbol{X}(t)\}_{t\geq 0} = \{(X_1(t), X_2(t))\}_{t\geq 0}$ be a time-inhomogeneous Markov jump process on E with transition intensity matrix function $\boldsymbol{M}$ and initial distribution $(\boldsymbol{\pi}_1(0), \boldsymbol{0})$. Let $\mathcal{S}_n = (T_i, Y_i)_{i\leq n}$ be the first n jump times and transitions of the macrostate process X_1. Then, for $\ell \in \{1, \dots, n\}$, the conditional process within a macro sojourn,*

$$W_\ell(t) \overset{\mathrm{d}}{=} \boldsymbol{X}(t) \big| \mathcal{S}_n \quad on \quad (T_{\ell-1} \leq t < T_\ell),$$

is a time-inhomogeneous Markov jump process taking values on $\{1, \dots, d_{Y_{\ell-1}}\}$ with initial distribution

$$\widetilde{\pi}_i^\ell(\mathcal{S}_n) = \frac{\boldsymbol{\alpha}(\mathcal{S}_{\ell-1})\boldsymbol{e}_i\boldsymbol{e}_i'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(T_{\ell-1}, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{\alpha}(\mathcal{S}_n)\boldsymbol{1}_{d_{Y_n}}},$$

transition probabilities

$$\widetilde{p}_{ij}^\ell(t, s \big| \mathcal{S}_n) = \frac{\boldsymbol{e}_i'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, s)\boldsymbol{e}_j\boldsymbol{e}_j'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(s, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{e}_i'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)},$$

and transition intensities

$$\widetilde{\mu}_{ij}^\ell(t \big| \mathcal{S}_n) = \mu_{(Y_{\ell-1},i)(Y_{\ell-1},j)}(t)\frac{\boldsymbol{e}_j'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{e}_i'\bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)},$$

where $\boldsymbol{e}_k$*,* $k \in \{1, \dots, d_{Y_{\ell-1}}\}$*, is a* $d_{Y_{\ell-1}}$*-dimensional column vector with one in entry* k *and zeros otherwise.*

Proof. It follows from Ahmad, Bladt, and Furrer (2022, Lemma A.1) that

$$\mathbb{P}(T_\ell \in \mathrm{d}t_\ell, Y_\ell = y_\ell \,|\, \mathcal{S}_{\ell-1} = \mathcal{s}_{\ell-1}) = \frac{\boldsymbol{\alpha}(\mathcal{s}_\ell)\mathbf{1}_{d_{y_\ell}}}{\boldsymbol{\alpha}(\mathcal{s}_{\ell-1})\mathbf{1}_{d_{y_{\ell-1}}}}\mathrm{d}t_\ell,$$

which implies, using that $\mathbb{P}(\mathcal{S}_0 = \mathcal{s}_0) = 1$,

$$\begin{aligned}
&\mathbb{P}(T_n \in \mathrm{d}t_n, Y_n = y_n, \dots, T_1 \in \mathrm{d}t_1, Y_1 = y_1) \\
&= \prod_{\ell=1}^{n} \mathbb{P}(T_\ell \in \mathrm{d}t_\ell, Y_\ell = y_\ell \,|\, \mathcal{S}_{\ell-1} = \mathcal{s}_{\ell-1}) &\quad (4.A.1)\\
&= \prod_{\ell=1}^{n} \frac{\boldsymbol{\alpha}(\mathcal{s}_\ell)\mathbf{1}_{d_{y_\ell}}}{\boldsymbol{\alpha}(\mathcal{s}_{\ell-1})\mathbf{1}_{d_{y_{\ell-1}}}}\mathrm{d}t_\ell. &\quad (4.A.2)\\
&= \boldsymbol{\alpha}(\mathcal{s}_n)\mathbf{1}_{d_{y_n}}\mathrm{d}t_1\,\mathrm{d}t_2 \cdots \mathrm{d}t_n.
\end{aligned}$$

The remaining part of the proof now largely follows the approach taken in Ahmad, Bladt, and Bladt (2022, Lemma 3.2). Fix $\ell \in \{1, \dots, n\}$ and $t, s \geq 0$. Applying same techniques as in (4.A.2), we get from the Markov property of $\boldsymbol{X}$, that on the event $(T_{\ell-1} \leq s < T_\ell)$,

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}_{(T_\ell \in \mathrm{d}t_\ell, Y_\ell = y_\ell, \dots, T_n \in \mathrm{d}t_n, Y_n = y_n)} \,\big|\, \mathcal{F}^{\boldsymbol{X}}(s)\big] \\
&= \boldsymbol{e}'_{X_2(s)}\bar{\boldsymbol{P}}_{X_1(s)}(s, t_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{s}_n)\,\mathrm{d}t_\ell \cdots \mathrm{d}t_n,
\end{aligned} \quad (4.A.3)$$

and similarly when conditioning on $\mathcal{F}^{\boldsymbol{X}}(t)$, on the corresponding event at time t. The transition probabilities for W_ℓ are then obtained as follows, on the event $(T_{\ell-1} \leq t < T_\ell)$,

$$\begin{aligned}
\mathbb{E}\big[\mathbb{1}_{(W_\ell(s) = \boldsymbol{j})} \,\big|\, \mathcal{F}^{W_\ell}(t)\big] &= \mathbb{E}\big[\mathbb{1}_{(\boldsymbol{X}(s) = \boldsymbol{j})}\mathbb{1}_{[T_{\ell-1}, T_\ell)}(s) \,\big|\, \mathcal{F}^{\boldsymbol{X}}(t) \vee \sigma(\mathcal{S}_n)\big] \\
&= \mathbb{E}\big[\mathbb{1}_{(\boldsymbol{X}(s) = \boldsymbol{j})}\mathbb{1}_{[T_{\ell-1}, T_\ell)}(s) \,\big|\, \mathcal{F}^{\boldsymbol{X}}(t) \vee \sigma(T_\ell, Y_\ell, \dots, T_n, Y_n)\big] \\
&= \frac{\mathbb{E}\big[\mathbb{1}_{(\boldsymbol{X}(s) = \boldsymbol{j})}\mathbb{1}_{[T_{\ell-1}, T_\ell)}(s) \,\big|\, \mathcal{F}^{\boldsymbol{X}}(t)\big]\, \boldsymbol{e}'_{\tilde{j}}\bar{\boldsymbol{P}}_j(s, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{e}'_{X_2(t)}\bar{\boldsymbol{P}}_{X_1(t)}(t, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)} \\
&= \frac{\boldsymbol{e}'_{X_2(t)}\bar{\boldsymbol{P}}_{X_1(t)}(t, s)\boldsymbol{e}_{\tilde{j}}\boldsymbol{e}'_{\tilde{j}}\bar{\boldsymbol{P}}_j(s, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{e}'_{X_2(t)}\bar{\boldsymbol{P}}_{X_1(t)}(t, T_\ell)\boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)},
\end{aligned}$$

where use the Markov property of $\boldsymbol{X}$ in the second equality, and the tower property with the sigma-algebra $\mathcal{F}^{\boldsymbol{X}}(s) \vee \sigma(T_\ell, Y_\ell, \dots, T_n, Y_n) \supseteq \mathcal{F}^{\boldsymbol{X}}(t) \vee \sigma(T_\ell, Y_\ell, \dots, T_n, Y_n)$ in the third equality. The fourth equality follows from Ahmad, Bladt, and Furrer

(2022, Lemma A.1 and Proposition 4.1). Conditioning on $\boldsymbol{X}(t) = \boldsymbol{i}$, we get the desired transition probabilities. From these, the corresponding transition intensities follow immediately:

$$
\begin{aligned}
\widetilde{\mu}^{\ell}_{ij}(t \big| \mathcal{S}_n) &= \lim_{h \downarrow 0} \frac{\widetilde{p}^{\ell}_{ij}(t, t+h \big| \mathcal{S}_n)}{h} \\
&= \frac{1}{\boldsymbol{e}_i' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)} \times \\
&\quad \lim_{h \downarrow 0} \frac{\boldsymbol{e}_i' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, t+h) \boldsymbol{e}_j}{h} \boldsymbol{e}_j' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(t+h, T_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n) \\
&= \mu_{(Y_{\ell-1}, i)(Y_{\ell-1}, j)}(t) \frac{\boldsymbol{e}_j' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{e}_i' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(t, T_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}.
\end{aligned}
$$

Here, we use the continuity of the transition (sub-)probability matrix function $\bar{\boldsymbol{P}}_{Y_{\ell-1}}$ obtained from the continuity of product integrals (whenever they exist). Lastly, to derive the initial distribution, we see using the same techniques as in (4.A.2), that

$$
\begin{aligned}
&\mathbb{P}\big(T_n \in \mathrm{d}t_n, Y_n = y_n, \ldots, T_{\ell-1} \in \mathrm{d}t_{\ell-1}, \\
&\qquad \boldsymbol{X}(T_{\ell-1}) = (y_{\ell-1}, \tilde{i}), \ldots, T_1 \in \mathrm{d}t_1, Y_1 = y_1\big) \qquad (4.A.4) \\
&= \boldsymbol{\alpha}(\mathfrak{s}_{\ell-1}) \boldsymbol{e}_{\tilde{i}} \boldsymbol{e}_{\tilde{i}}' \bar{\boldsymbol{P}}_{y_{\ell-1}}(t_{\ell-1}, t_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathfrak{s}_n)\, \mathrm{d}t_1 \cdots \mathrm{d}t_n, \qquad (4.A.5)
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathbb{P}(W_\ell(T_{\ell-1}) = \boldsymbol{i}) &= \mathbb{E}\big[\mathbb{1}_{(\boldsymbol{X}(T_{\ell-1}) = \boldsymbol{i})} \,\big|\, \sigma(\mathcal{S}_n)\big] \\
&= \frac{\boldsymbol{\alpha}(\mathcal{S}_{\ell-1}) \boldsymbol{e}_{\tilde{i}} \boldsymbol{e}_{\tilde{i}}' \bar{\boldsymbol{P}}_{Y_{\ell-1}}(T_{\ell-1}, T_\ell) \boldsymbol{\alpha}_{\ell+1}(\mathcal{S}_n)}{\boldsymbol{\alpha}(\mathcal{S}_n) \mathbf{1}_{d_{y_n}}},
\end{aligned}
$$

as desired. □

Proof of Theorem 4.4.1. The proof largely follows the approach taken in Ahmad, Bladt, and Bladt (2022, Theorem 3.4). For notational convenience, we write $\mathbb{E}^{(m)}$ for the expectation operator $\mathbb{E}_{(\boldsymbol{\pi}_1^{(m)}, \boldsymbol{\theta}^{(m)})}$. Now, it is evident from an application of Fubini's theorem that the conditional expected log-likelihood (4.4.1) is on the form (4.4.7), and it, therefore, suffices to show the results for the conditional expected statistics (4.4.4)-(4.4.6). By independence between the data points in $\boldsymbol{\mathcal{S}}$ and Lemma 4.A.1, we get

$$
\bar{B}^{(m)}_{(1,r)}(0) = \sum_{n=1}^{N} \mathbb{E}^{(m)}\Big[\mathbb{1}_{(\boldsymbol{X}^{(n)}(0) = (1,r))} \,\big|\, \mathcal{S}^{(n)}\Big] = \tilde{\pi}^0_r\big(\mathcal{S}^{(n)}\big),
$$

which by insertion yields the desired result. For $\bar{I}^{(m)}_{\boldsymbol{i}}$, we get

$$
\bar{I}^{(m)}_{\boldsymbol{i}}(u) = \sum_{n=1}^{N} \mathbb{E}^{(m)}\Big[\mathbb{1}_{(\boldsymbol{X}^{(n)}(u) = \boldsymbol{i})} \,\big|\, \mathcal{S}^{(n)}\Big]
$$

$$= \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)}(u) \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \sum_{\check{i}=1}^{d_i} \widetilde{\pi}_{\check{i}}^{\ell}\big(\mathcal{S}^{(n)}\big) \widetilde{p}_{\check{i}\,\tilde{i}}\big(T_{\ell-1}^{(n)}, u \vert \mathcal{S}^{(n)}\big).$$

The inner sum is then given by

$$\sum_{\check{i}=1}^{d_i} \widetilde{\pi}_{\check{i}}^{\ell}\big(\mathcal{S}^{(n)}\big) \widetilde{p}_{\check{i}\tilde{i}}\big(T_{\ell-1}^{(n)}, u \vert \mathcal{S}^{(n)}\big)$$

$$= \sum_{\check{i}=1}^{d_i} \frac{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}_{\ell-1}^{(n)}\big) \boldsymbol{e}_{\check{i}} \boldsymbol{e}_{\check{i}}' \bar{\boldsymbol{P}}_i^{(m)}\big(T_{\ell-1}^{(n)}, u\big) \boldsymbol{e}_{\tilde{i}} \boldsymbol{e}_{\tilde{i}}' \bar{\boldsymbol{P}}_i^{(m)}\big(u, T_\ell^{(n)}\big) \boldsymbol{\alpha}_{\ell+1}^{(m)}\big(\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big) \mathbf{1}_{d_{y_n}}}$$

$$= \frac{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}_{\ell-1}^{(n)}\big) \bar{\boldsymbol{P}}_i^{(m)}\big(T_{\ell-1}^{(n)}, u\big) \boldsymbol{e}_{\tilde{i}} \boldsymbol{e}_{\tilde{i}}' \bar{\boldsymbol{P}}_i^{(m)}\big(u, T_\ell^{(n)}\big) \boldsymbol{\alpha}_{\ell+1}^{(m)}\big(\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big) \mathbf{1}_{d_{y_n}}},$$

which shows the result for $\bar{I}_{\boldsymbol{i}}^{(m)}$. For $\bar{N}_{\boldsymbol{i}(i,\check{i})}^{(m)}$, $\check{i} \in \{1, \ldots, d_i\}$, $\check{i} \neq \tilde{i}$, we get

$$\bar{N}_{\boldsymbol{i}(i,\check{i})}^{(m)}(u) = \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mathbb{E}^{(m)}\left[\int_{(0,u]} \mathbb{1}_{\left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)}(x) \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \, \mathrm{d}N_{\boldsymbol{i}(i,\check{i})}^{(n)}(x) \,\middle|\, \mathcal{S}^{(n)} \right].$$

Then, using the intensity process of $\{\boldsymbol{X}^{(n)}(x)\}_{x \in \left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)} \big\vert \mathcal{S}^{(n)}$ from Lemma 4.A.1,

$$\begin{aligned}
&\bar{N}_{\boldsymbol{i}(i,\check{i})}^{(m)}(u) \\
&= \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mathbb{E}^{(m)}\left[\int_0^u \mathbb{1}_{\left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)}(x) \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \mathbb{1}_{(\boldsymbol{X}^{(n)}(x)=\boldsymbol{i})} \widetilde{\mu}_{\tilde{i}\check{i}}^{\ell}\big(x \big\vert \mathcal{S}^{(n)}\big) \mathrm{d}x \,\middle|\, \mathcal{S}^{(n)} \right] \\
&= \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \int_0^u \mathbb{1}_{\left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)}(x) \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \times \\
&\qquad \sum_{r=1}^{d_i} \widetilde{\pi}_r^{\ell}\big(\mathcal{S}^{(n)}\big) \widetilde{p}_{r\tilde{i}}\big(T_{\ell-1}^{(n)}, x \vert \mathcal{S}^{(n)}\big) \widetilde{\mu}_{\tilde{i}\check{i}}^{\ell}\big(x \big\vert \mathcal{S}^{(n)}\big) \, \mathrm{d}x \\
&= \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \int_0^u \mathbb{1}_{\left[T_{\ell-1}^{(n)}, T_\ell^{(n)}\right)}(x) \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \times \\
&\quad \frac{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}_{\ell-1}^{(n)}\big) \bar{\boldsymbol{P}}_i^{(m)}\big(T_{\ell-1}^{(n)}, x\big) \boldsymbol{e}_{\tilde{i}} \mu_{\boldsymbol{i}(i,\check{i})}\big(x; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_{\check{i}}' \bar{\boldsymbol{P}}_i^{(m)}\big(x, T_\ell^{(n)}\big) \boldsymbol{\alpha}_{\ell+1}^{(m)}\big(\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big) \mathbf{1}_{d_{y_n}}} \, \mathrm{d}x.
\end{aligned}$$

Taking dynamics in u then yields the desired result. Finally, for $\bar{N}_{\boldsymbol{i}\boldsymbol{j}}^{(m)}$, $j \neq i$, we see that $N_{\boldsymbol{i}\boldsymbol{j}}$ can be written as

$$N_{\boldsymbol{i}\boldsymbol{j}}(u) = \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i,\, Y_\ell^{(n)}=j\right)} \mathbb{1}_{\left(T_\ell^{(n)} \leq u\right)} \mathbb{1}_{\left(\boldsymbol{X}^{(n)}(T_\ell^{(n)}-)=\boldsymbol{i}\right)} \mathbb{1}_{\left(\boldsymbol{X}^{(n)}(T_\ell^{(n)})=\boldsymbol{j}\right)}$$

such that

$$\bar{N}_{\boldsymbol{\imath}\boldsymbol{\jmath}}^{(m)}(u) = \sum_{n=1}^{N}\sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i,\, Y_{\ell}^{(n)}=j\right)} \mathbb{1}_{\left(T_\ell^{(n)} \leq u\right)} \times \\ \mathbb{E}^{(m)}\left[\mathbb{1}_{\left(\boldsymbol{X}^{(n)}(T_\ell^{(n)}-)=\boldsymbol{\imath}\right)} \mathbb{1}_{\left(\boldsymbol{X}^{(n)}(T_\ell^{(n)})=\boldsymbol{\jmath}\right)} \,\middle|\, \sigma\big(\mathcal{S}^{(n)}\big)\right].$$

Now, since

$$\mathbb{P}^{(m)}\Big(\boldsymbol{X}^{(n)}(T_\ell^{(n)}) = \boldsymbol{\jmath} \,\Big|\, \boldsymbol{X}^{(n)}(T_\ell^{(n)}-) = \boldsymbol{\imath} \Big) = \mu_{\boldsymbol{\imath}\boldsymbol{\jmath}}\big(T_\ell^{(n)}; \boldsymbol{\theta}^{(m)}\big),$$

we get, using the same techniques as in (4.A.2)-(4.A.5), that

$$\begin{aligned}
&\mathbb{P}\Big(T_{M^{(n)}}^{(n)} \in \mathrm{d}t_{m^{(n)}}^{(n)}, Y_{M^{(n)}}^{(n)} = y_{m^{(n)}}^{(n)}, \dots, T_\ell^{(n)} \in \mathrm{d}t_\ell^{(n)}, \boldsymbol{X}(T_\ell^{(n)}) = \boldsymbol{\jmath}, \boldsymbol{X}(T_\ell^{(n)}-) = \boldsymbol{\imath}, \\
&\quad T_{\ell-1}^{(n)} \in \mathrm{d}t_{\ell-1}^{(n)}, Y_{\ell-1}^{(n)} = i, \dots, T_1 \in \mathrm{d}t_1, Y_1 = y_1 \Big) \\
&= \boldsymbol{\alpha}^{(m)}\big(\mathfrak{s}_{\ell-1}^{(n)}\big) \bar{\boldsymbol{P}}_i^{(m)}\big(t_{\ell-1}^{(n)}, t_\ell^{(n)}\big) \boldsymbol{e}_{\tilde{\imath}} \mu_{\boldsymbol{\imath}\boldsymbol{\jmath}}\big(t_\ell^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_{\tilde{\jmath}}' \bar{\boldsymbol{P}}_j^{(m)}\big(t_\ell^{(n)}, t_{\ell+1}^{(n)}\big) \boldsymbol{\alpha}_{\ell+2}^{(m)}\big(\mathfrak{s}^{(n)}\big) \times \\
&\quad \mathrm{d}t_1^{(n)} \cdots \mathrm{d}t_{m^{(n)}}^{(n)}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\bar{N}_{\boldsymbol{\imath}\boldsymbol{\jmath}}^{(m)}(u) \\
&= \sum_{n=1}^{N}\sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i,\, Y_{\ell}^{(n)}=j\right)} \mathbb{1}_{\left(T_\ell^{(n)} \leq u\right)} \times \\
&\frac{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}_{\ell-1}^{(n)}\big) \bar{\boldsymbol{P}}_i^{(m)}\big(T_{\ell-1}^{(n)}, T_\ell^{(n)}\big) \boldsymbol{e}_{\tilde{\imath}} \mu_{\boldsymbol{\imath}\boldsymbol{\jmath}}\big(T_\ell^{(n)}; \boldsymbol{\theta}^{(m)}\big) \boldsymbol{e}_{\tilde{\jmath}}' \bar{\boldsymbol{P}}_j^{(m)}\big(T_\ell^{(n)}, T_{\ell+1}^{(n)}\big) \boldsymbol{\alpha}_{\ell+2}^{(m)}\big(\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n}.
\end{aligned}$$

Taking dynamics in u then yields the desired result. □

Proof of Corollary 4.4.2. It is evident from the complete data log-likelihood (4.3.10) that the conditional expectation (4.4.8) is on the form (4.4.11), by an application of Fubini's theorem. It, therefore, suffices to show that the conditional statistics are given as in (4.4.13)-(4.4.14). To obtain these, the defining property is, as also noted in Ahmad, Bladt, and Furrer (2022, Subsection 4.2), that the reset property (4.2.6) implies that for $j \neq i$,

$$\boldsymbol{\pi}_i\big(t; \boldsymbol{\eta}^{(m)}\big) \bar{\boldsymbol{P}}_i^{(m)}(t,s) \boldsymbol{\beta}_{ij}\big(s; \boldsymbol{\theta}^{(m)}\big)$$

is a 1×1-dimensional matrix, and thus cancels if appearing in both the numerator and denominator of a fraction. Concerning the conditional statistics within macrostates,

$\bar{I}_{\boldsymbol{i}}^{(m)}$ and $\bar{N}_{\boldsymbol{i}(i,\check{i})}^{(m)}$, this property implies

$$\sum_{n=1}^{N} \frac{\boldsymbol{c}_i^{(m)}\big(u;\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n} = \sum_{n=1}^{N}\sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left[T_{\ell-1}^{(n)},\,T_\ell^{(n)}\right)}(u)\mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i\right)} \frac{\boldsymbol{c}_i^{(m)}\big(u;T_{\ell-1}^{(n)},i,T_\ell^{(n)},Y_\ell^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(T_{\ell-1}^{(n)},i,T_\ell^{(n)},Y_\ell^{(n)}\big)\mathbf{1}_n}, \tag{4.A.6}$$

where, for generic values $0 \le x_1 < x_2 < \infty$, and $j \in \mathcal{J}$, $j \neq i$, we have for $u \in [x_1, x_2)$,

$$\begin{aligned} \boldsymbol{c}_i^{(m)}(u;x_1,i,x_2,j) &= \bar{\boldsymbol{P}}_i^{(m)}(u,x_2)\boldsymbol{\beta}_{ij}\big(x_2;\boldsymbol{\theta}^{(m)}\big)\boldsymbol{\pi}_i\big(x_1;\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}(x_1,u), \\ \boldsymbol{\alpha}^{(m)}(x_1,i,x_2,j)\mathbf{1}_{d_j} &= \boldsymbol{\pi}_i\big(x_1;\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}(x_1,x_2)\boldsymbol{\beta}_{ij}\big(x_2;\boldsymbol{\theta}^{(m)}\big), \end{aligned} \tag{4.A.7}$$

and zero otherwise. Hence, for a fixed time u between two jump times, each term on the right-hand side of (4.A.6) only depends on the last jump before u, the next jump time after u, and state jumped to at the next jump time; the past and future sojourns outside time u are cancelled out. Consequently, when summing over all observations and corresponding sojourns, one can equivalently sum over the data points in $\mathcal{T}_i$, as these provide the jump times and states needed (cf. (4.4.9)). This gives

$$\sum_{n=1}^{N} \frac{\boldsymbol{c}_i^{(m)}\big(u;\mathcal{S}^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n} = \sum_{n=1}^{M_i} \frac{\boldsymbol{c}_i^{(m)}\big(u;\mathcal{T}_i^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}_i^{(n)}\big)\mathbf{1}_n}, \tag{4.A.8}$$

which shows (4.4.13) for $\bar{I}_{\boldsymbol{i}}^{(m)}$ and $\bar{N}_{\boldsymbol{i}(i,\check{i})}^{(m)}$. Concerning the conditional statistics for jumps between macrostates, $\bar{N}_{\boldsymbol{i}j}^{(m)}$, we use $\bar{N}_{\boldsymbol{i}\boldsymbol{j}}^{(m)}$ from Theorem 4.4.1 to get

$$\begin{aligned} \bar{N}_{\boldsymbol{i}j}^{(m)}(u) &= \sum_{\tilde{j}=1}^{d_j} \bar{N}_{\boldsymbol{i}\boldsymbol{j}}(u) \\ &= \sum_{n=1}^{N}\sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left(T_\ell^{(n)}\ge u\right)}\beta_{\boldsymbol{i}j}\big(T_\ell^{(n)};\boldsymbol{\theta}^{(m)}\big)\boldsymbol{\pi}_j\big(T_\ell^{(n)};\boldsymbol{\eta}^{(m)}\big)\frac{\boldsymbol{a}_{ij}^{(m)}\big(T_\ell^{(n)};\mathcal{S}^{(n)}\big)\boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{S}^{(n)}\big)\mathbf{1}_n}. \end{aligned}$$

Using the same technique with the fraction as in (4.A.6)-(4.A.8), we get

$$\begin{aligned} &\bar{N}_{\boldsymbol{i}j}^{(m)}(u) \\ &= \sum_{n=1}^{N}\sum_{\ell=1}^{M^{(n)}} \mathbb{1}_{\left(T_\ell^{(n)}\ge u\right)}\mathbb{1}_{\left(Y_{\ell-1}^{(n)}=i,\,Y_\ell^{(n)}=j\right)}\times \\ &\qquad \frac{\boldsymbol{\pi}_i\big(T_{\ell-1}^{(n)};\boldsymbol{\eta}^{(m)}\big)\bar{\boldsymbol{P}}_i^{(m)}\big(T_{\ell-1}^{(n)},T_\ell^{(n)}\big)\boldsymbol{e}_{\tilde{i}}\beta_{\boldsymbol{i}j}\big(T_\ell^{(n)};\boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\alpha}^{(m)}\big(T_{\ell-1}^{(n)},i,T_\ell^{(n)},j\big)\mathbf{1}_n} \end{aligned}$$

$$
= \sum_{n=1}^{M_i} \mathbb{1}\left(\tau_i^{(n)} \geq u\right) \mathbb{1}\left(Z_i^{(n)} = j\right) \frac{\boldsymbol{\pi}_i\big(R_i^{(n)}; \boldsymbol{\eta}^{(m)}\big) \bar{\boldsymbol{P}}_i^{(m)}\big(R_i^{(n)}, \tau_i^{(n)}\big) \boldsymbol{e}_{\tilde{i}} \beta_{\boldsymbol{i}j}\big(\tau_i^{(n)}; \boldsymbol{\theta}^{(m)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\mathcal{T}_i^{(n)}\big) \mathbf{1}_n}.
$$

Taking dynamics in u then yields (4.4.14). Concerning the conditional statistics for initiations into macrostates, $\bar{N}_{\boldsymbol{i}}^{(m)}$, we use the exact same approach as for $\bar{N}_{\boldsymbol{i}j}^{(m)}$, with the only difference being the summation over $\boldsymbol{j}$ of $\boldsymbol{e}_{\tilde{j}} \beta_{\boldsymbol{j}i}(\cdot; \boldsymbol{\theta}^{(m)})$ instead of a summation over $\tilde{j}$ of $\pi_{\boldsymbol{j}}(\cdot; \boldsymbol{\eta}^{(m)}) \boldsymbol{e}'_{\tilde{j}}$. □

4.B General EM algorithms

Algorithm 4.2 EM algorithm for the general aggregate Markov model

Input: Data points $\boldsymbol{\jmath} = (\jmath^{(1)}, \dots, \jmath^{(N)})$, each element consisting of $m^{(n)}$ jump times and transitions on the macro level, and initial parameters $(\boldsymbol{\pi}_1^{(0)}, \boldsymbol{\theta}^{(0)}) \in [0,1]^{d_1} \times \boldsymbol{\Theta}$.

0) Set $m := 0$

1) *E-step:* For $r \in \{1, \dots, d_1\}$, compute conditional statistics for the initial state,

$$\bar{B}^{(m)}_{(1,r)}(0) = \sum_{n=1}^{N} \frac{\pi^{(m)}_{(1,r)}(0)\, \boldsymbol{e}_r' \bar{\boldsymbol{P}}_1^{(m)}\big(0, t_1^{(n)}\big) \boldsymbol{\alpha}_1^{(m)}\big(\jmath^{(n)}\big)}{\boldsymbol{\alpha}^{(m)}\big(\jmath^{(n)}\big)\mathbf{1}_n},$$

and, for $\boldsymbol{i}, \boldsymbol{j} \in E$, $\boldsymbol{j} \neq \boldsymbol{i}$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (on a suitable grid), compute the conditional expected log-likelihood for the transitions:

$$\bar{L}^{(m)}_{\boldsymbol{i}\boldsymbol{j}}(\boldsymbol{\theta}) = \int_{(0,T]} \log(\mu_{\boldsymbol{i}\boldsymbol{j}}(u;\boldsymbol{\theta}))\, \mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}\boldsymbol{j}}(u) - \int_0^T \bar{I}^{(m)}_{\boldsymbol{i}}(u) \mu_{\boldsymbol{i}\boldsymbol{j}}(u;\boldsymbol{\theta})\, \mathrm{d}u,$$

where

$$\bar{I}^{(m)}_{\boldsymbol{i}}(u) = \sum_{n=1}^{N} \frac{\boldsymbol{e}_{\tilde{i}}' \boldsymbol{c}_i^{(m)}\big(u; \jmath^{(n)}\big) \boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\jmath^{(n)}\big)\mathbf{1}_n},$$

and, for $\check{i} \in \{1, \dots, d_i\}$, $\check{i} \neq \tilde{i}$,

$$\mathrm{d}\bar{N}_{\boldsymbol{i}(i,\check{i})}(u) = \sum_{n=1}^{N} \mu_{\boldsymbol{i}(i,\check{i})}\big(u; \boldsymbol{\theta}^{(m)}\big) \frac{\boldsymbol{e}_{\check{i}}' \boldsymbol{c}_i^{(m)}\big(u; \jmath^{(n)}\big) \boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\jmath^{(n)}\big)\mathbf{1}_n} \mathrm{d}u,$$

while for $j \neq i$,

$$\mathrm{d}\bar{N}^{(m)}_{\boldsymbol{i}\boldsymbol{j}}(u) = \sum_{n=1}^{N} \sum_{\ell=1}^{M^{(n)}} \mu_{\boldsymbol{i}\boldsymbol{j}}\big(u; \boldsymbol{\theta}^{(m)}\big) \frac{\boldsymbol{e}_{\tilde{j}}' \boldsymbol{a}_{ij}^{(m,\ell)}\big(u; \jmath^{(n)}\big) \boldsymbol{e}_{\tilde{i}}}{\boldsymbol{\alpha}^{(m)}\big(\jmath^{(n)}\big)\mathbf{1}_n}\, \mathrm{d}\varepsilon_{T_\ell^{(n)}}(u).$$

2) *M-step:* Update the parameters:

$$\hat{\pi}^{(m+1)}_{(1,r)}(0) = \frac{\bar{B}^{(m)}_{(1,r)}(0)}{N},$$

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in E \\ \boldsymbol{j} \neq \boldsymbol{i}}} \bar{L}^{(m)}_{\boldsymbol{i}\boldsymbol{j}}(\boldsymbol{\theta}).$$

3) Set $m := m + 1$ and GOTO 1) until a stopping rule is satisfied.

Output: *Fitted parameters* $(\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\theta}})$.

Algorithm 4.3 EM algorithm for the aggregate Markov model with the reset property

Input: Initial parameters $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\eta}^{(0)}) \in \boldsymbol{\Theta} \times \boldsymbol{H}$, and for each macrostate $i \in \mathcal{J}$, data points within the macrostate,

$$\boldsymbol{\mathcal{T}}_i = \left\{ \left(r_i^{(n)}, \tau_i^{(n)}, z_i^{(n)} \right) \right\}_{n \in \{1,\ldots,m_i\}},$$

consisting of time of entries $r_i^{(n)}$ into the macrostate, jump times $\tau_i^{(n)}$ out of the state, and macrostate $z_i^{(n)}$ jumped to at time $\tau_i^{(n)}$.

0) Set $m := 0$

1) *E-step:* For each macrostate $i \in \mathcal{J}$,

- For each $n \in \{1, \ldots, m_i^{(n)}\}$, run the E-step of the EM algorithm in Ahmad, Bladt, and Bladt (2022, Appendix A) with
  - Parameter space $\boldsymbol{\Theta}$
  - State-space of transient states $\{1, \ldots, d_i\}$
  - A single IPH observation $\tau_i^{(n)} - r_i^{(n)}$
  - Initial distribution $\boldsymbol{\pi}_i\big(r_i^{(n)}; \boldsymbol{\eta}^{(m)}\big)$
  - Sub-intensity matrix function $x \mapsto \boldsymbol{M}_{ii}\big(x + r_i^{(n)}; \boldsymbol{\theta}^{(m)}\big)$
  - Exit rate vector function $x \mapsto \boldsymbol{\beta}_{i z_i^{(n)}}\big(x + r_i^{(n)}; \boldsymbol{\theta}^{(m)}\big)$

  which outputs:
  - Conditional statistics for the initial state: $\bar{N}_{\boldsymbol{i}}^{(n,m)}\big(R_i^{(n)}\big)$
  - Likelihood for transitions between transient states: $\bar{L}_{\boldsymbol{i}(i,\check{i})}^{(n,m)}(\boldsymbol{\theta})$
  - Likelihood for transitions to the 'absorbing' state: $\bar{L}_{\boldsymbol{i} z_i^{(n)}}^{(n,m)}(\boldsymbol{\theta})$

  for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (on a suitable grid).

2) *M-step:* Update the parameters:

$$\hat{\boldsymbol{\eta}}^{(m+1)} = \arg\max_{\boldsymbol{\eta}} \sum_{\boldsymbol{i} \in E} \sum_{n=1}^{m_i} \bar{N}_{\boldsymbol{i}}^{(n,m)}\big(R_i^{(n)}\big) \log\big(\pi_{\boldsymbol{i}}\big(R_i^{(n)}; \boldsymbol{\eta}\big)\big)$$

$$\hat{\boldsymbol{\theta}}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\boldsymbol{i} \in E} \sum_{n=1}^{m_i} \left(\sum_{\substack{\check{i}=1 \\ \check{i} \neq i}}^{d_i} \bar{L}_{\boldsymbol{i}(i,\check{i})}^{(n,m)}(\boldsymbol{\theta}) + \sum_{\substack{j \in \mathcal{J} \\ j \neq i}} \bar{L}_{\boldsymbol{i}j}^{(n,m)}(\boldsymbol{\theta}) \right).$$

3) Set $m := m + 1$ and GOTO 1) until a stopping rule is satisfied.

Output: *Fitted parameters* $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$.

Chapter 5

Phase-type representations of stochastic interest rates with applications to life insurance

This chapter is based on the manuscript *Ahmad and Bladt (2022b).*

Abstract

The purpose of the present paper is to incorporate stochastic interest rates into a matrix-approach to multi-state life insurance, where formulas for reserves, moments of future payments and equivalence premiums can be obtained as explicit formulas in terms of product integrals or matrix exponentials. To this end we consider the Markovian interest model, where the rates are piecewise deterministic (or even constant) in the different states of a Markov jump process, and which is shown to integrate naturally into the matrix framework. The discounting factor then becomes the price of a zero-coupon bond which may or may not be correlated with the biometric insurance process. Another nice feature about the Markovian interest model is that the price of the bond coincides with the survival function of a phase-type distributed random variable. This, in particular, allows for calibrating the Markovian interest rate models using a maximum likelihood approach to observed data (prices) or to theoretical models like e.g. a Vasiček model. Due to the denseness of phase-type distributions, we can approximate the price behavior of any zero-coupon bond with interest rates bounded from below by choosing the number of possible interest rate values sufficiently large. For observed data models with few data points, lower dimensions will usually suffice, while for theoretical models the dimensionality is only a computational issue.

Keywords: Zero-coupon bond; Phase-type distribution; Stochastic interest rate; Multi-state life insurance; Thiele's differential equation

5.1 Introduction

This paper considers stochastic interest models, which are state-wise deterministic dependent on an underlying finite state-space Markov process. The spot rate $r(u)$ at time u is assumed to be on the form

$$r(u) = r_{X(u)}(u), \tag{5.1.1}$$

where $\{X(u)\}_{u\geq 0}$ denotes a time-inhomogeneous Markov jump process on a p-dimensional state-space, and $r_i(u)$, $i = 1, ..., p$, are deterministic functions. Assuming an arbitrage free bond market, a zero-coupon bond with terminal date T can then be defined in terms of its prices by

$$B(t,T) = \mathrm{E}^{\mathbb{Q}}\Big(\mathrm{e}^{-\int_t^T r_{X(u)}(u)\,\mathrm{d}u} \,\Big|\, \mathcal{F}(t)\Big), \quad 0 \leq t \leq T, \tag{5.1.2}$$

where $\mathcal{F}(t) = \sigma(X(u) : 0 \leq u \leq t)$ is the σ-algebra generated by $\{X(u)\}_{u\geq 0}$. The expectation is taken under some risk-neutral measure $\mathbb{Q}$ (see, e.g., Björk, 2009; Elliott and Kopp, 1999). If all $r_i(u) \geq 0$, a key result of the paper is that, conditionally on $X(t)$, $T \to B(t,T)$ equals the survival function of an inhomogeneous phase-type distribution.

In the presence of negative interest rates, this is longer certain since $B(t,T)$ may be larger than one and non-monotone. However, assuming that the negative interest rates are bounded from below by a number $-\rho < 0$, we get from (5.1.2) that

$$\mathrm{e}^{-\rho(T-t)}B(t,T) = \mathrm{E}^{\mathbb{Q}}\Big(\mathrm{e}^{-\int_t^T (r_{X(u)}(u)+\rho)\,\mathrm{d}u} \,\Big|\, \mathcal{F}(t)\Big) \tag{5.1.3}$$

then equals a survival function of an inhomogeneous phase-type distribution.

The interpretation that the bond prices are (possibly scaled) phase-type survival functions enables us to fit (calibrate) the transition rates of $\{X(u)\}_{u\geq 0}$ from the observed bond prices by using a maximum likelihood approach. Since phase-type distributions are dense, i.e. can approximate any distribution with a sufficient number of phases, we may then fit a PH to the observed survival function (equivalent to a histogram) such that all observations (bond prices) are hit. The last point of observation may be considered right censored. All fitted transition rates are under a risk-neutral measure $\mathbb{Q}$.

The functional form of the state-wise price of the bond was noted already in Norberg (2003, (3.17)), though its relation to phase-type theory was not mentioned, and its potential was not further explored. We also believe that the "bond price representation" (5.1.2) of a phase-type survival function is unknown to the phase-type community.

In the context of multi-state life insurance, modeling stochastic interest rates also play a crucial role. The literature varies from SDE based models, see e.g. Norberg and Møller (1996), Møller and Steffensen (2007), Buchardt (2014), Baños (2020), and Asmussen and Steffensen (2020), to the finite state-space Markov chain models of Norberg (1995a,b) on the form (5.1.1). In the SDE-based methods, one often relies on an independence assumption between interest rates and biometric risk so that available forward rate curves can be used for valuation; an exception is Buchardt (2014), where dependence between interest rates and biometric risk is incorporated. In either case, the SDE-based models do not integrate into classic Thiele and Hattendorff type of results, which limits time-dynamic valuations based on these traditional methods.

The spot rate model (5.1.1), however, can be wholly incorporated into Thiele and Hattendorff type of differential equations for reserves and higher order moments, as shown by Norberg (1995b,a) and further explored in Norberg (2003). These observations allow for dependency between interest rates and transitions in life insurance, as well as time-dynamic valuations, without altering the traditional methods. The latter refers to the model (5.1.1) as the Markov chain market while Koller (2012) refers to it as Markovian interest intensities.

In this paper, we work with an extended version of the bond prices,

$$\mathrm{E}^{\mathbb{Q}}\Big(\mathbb{1}_{(X(T)=j)}\mathrm{e}^{-\int_t^T r_{X(u)}(u)\,\mathrm{d}u}\,\Big|\,\mathcal{F}(t)\Big), \quad j=1,...,p, \tag{5.1.4}$$

which in an insurance context are the discounting factors on the event that the terminal state will be j. Providing a matrix-representation for (5.1.4), we then find how it naturally integrates into the matrix framework of Bladt, Asmussen, and Steffensen (2020). The extension is convenient from a mathematical point of view and also relates to the partial (Bladt, Asmussen, and Steffensen, 2020) and retrospective reserves in single states (Subsection 5.E in Norberg, 1991). The treatment of the latter, however, is outside the scope of the current paper. We restate the results of the former framework in the context of stochastic interest rates. The proofs, and parts of the exposition, will differ from that of Bladt, Asmussen, and Steffensen (2020).

Markov jump processes in finance are often used in connection with regime switching models or where the different states are used to alter the parameters of usually SDE-driven processes. Here transitions can take place under some physical measure and may have a real-world interpretation. The Markov chain model for interest rates (5.1.1) can be thought of as a regime-switching model under a risk-neutral measure, particularly if the interest rates for each state are known a priori.

The Markov jump process approach can approximate bond price modeling in terms of diffusions. Formal constructions have been made in Bharucha-Reid (1960), Kurtz (1970, 1978), and Mijatović and Pistorius (2013). Since phase-type distributions form a dense class of distributions on the positive reals, this paper will offer an alternative and parsimonious way to approximate any zero-coupon bond (arbitrarily close) by a bond on the form (5.1.2).

The paper is organised as follows. Section 5.2 introduces some background and notation. Bond price modeling using phase-type distribution is developed in Section 5.3. In Section 5.4, we develop estimation of the Markovian interest rate model, both with and without restricted interest rates, and we provide examples of calibration to diffusion models and real data. In Section 5.5 we adjust the life insurance framework of Bladt, Asmussen, and Steffensen (2020) to allow for stochastic interest rates of the form (5.1.1). It contains examples of how to set up a model using the fitted bond parameters of Section 5.3 as well as a matrix-based method for calculating the equivalence premium, either via Newton's method or as an explicit formula. In Section 5.6 we present a numerical example. For the sake of exposition, the proofs are deferred to Appendix 5.B.

5.2 Background

5.2.1 Notation

Unless otherwise stated, row vectors are denoted by bold Greek lowercase letters (e.g., $\boldsymbol{\pi}$) and column vectors by bold lowercase Roman letters (e.g., $\boldsymbol{v}$). Elements of vectors are denoted by the same unbold, indexed letters (like $\boldsymbol{v} = (v_1, ..., v_p)'$). The vector $\boldsymbol{e}_i$ is the column vector which is 1 at index i and zero otherwise whereas $\boldsymbol{e} = (1, 1, ..., 1)'$.

Matrices are denoted by bold capital letters (Greek or Roman) and their elements by their corresponding lowercase indexed letters (e.g. $\boldsymbol{A} = \{a_{ij}\}$). If $\boldsymbol{v}$ is a vector (row or column), then $\boldsymbol{\Delta}(\boldsymbol{v})$ denotes the diagonal matrix, which has $\boldsymbol{v}$ as diagonal.

5.2.2 The product integral

Consider a time-inhomogeneous Markov jump process $X = \{X(t)\}_{t\geq 0}$ taking values in a finite state space $E = \{1, \dots, p\}$, with intensity matrix (functions) $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j\in E}$. Denote by $\boldsymbol{P}(s,t) = \{p_{ij}(s,t)\}$ the corresponding transition matrix, the elements of which are the transition probabilities $p_{ij}(s,t) = \mathbb{P}(X(t) = j|X(s) = i)$ for $i, j \in E$. The transition matrix $\boldsymbol{P}(s,t)$ then satisfies Kolmogorov's forward and backward differential equations,

$$
\begin{aligned}
&\frac{\partial}{\partial t}\boldsymbol{P}(s,t) = \boldsymbol{P}(s,t)\boldsymbol{M}(t), \qquad \boldsymbol{P}(s,s) = \boldsymbol{I}, \\
&\frac{\partial}{\partial s}\boldsymbol{P}(s,t) = -\boldsymbol{M}(s)\boldsymbol{P}(s,t), \quad \boldsymbol{P}(t,t) = \boldsymbol{I}.
\end{aligned}
\tag{5.2.1}
$$

The solution to (5.2.1), which in general is not explicitly available, will be denoted by

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x) \tag{5.2.2}$$

and referred to as the product integral of $\boldsymbol{M}(x)$ from s to t. This is also true for general matrix functions $\boldsymbol{M}(t)$, which satisfy (5.2.1) but are not intensity matrices.

Product integrals have several nice properties. For any $s, t, u \geq 0$, it satisfies the product rule

$$\prod_s^u (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x) \prod_t^u (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x), \tag{5.2.3}$$

which in turn implies that the product integral is invertible with

$$\left[\prod_s^t (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x)\right]^{-1} = \prod_t^s (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x). \tag{5.2.4}$$

If all $\boldsymbol{M}(x)$ commute, then

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x) = \exp\left(\int_s^t \boldsymbol{M}(x)\,\mathrm{d}x\right). \tag{5.2.5}$$

In particular, for $\boldsymbol{M}(x) \equiv \boldsymbol{M}$, we get

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{M}(x)\,\mathrm{d}x) = \mathrm{e}^{\boldsymbol{M}(t-s)}. \tag{5.2.6}$$

If $\boldsymbol{A}(x)$ and $\boldsymbol{B}(y)$ commute for all x, y, then

$$\prod_s^t (\boldsymbol{I} + (\boldsymbol{A}(x) + \boldsymbol{B}(x))\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \prod_s^t (\boldsymbol{I} + \boldsymbol{B}(x)\,\mathrm{d}x). \tag{5.2.7}$$

In particular,

$$\mathrm{e}^{-r(t-s)} \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + [\boldsymbol{A}(x) - r\boldsymbol{I}]\,\mathrm{d}x), \tag{5.2.8}$$

where $\boldsymbol{I}$ denotes the identity matrix.

Remark 5.2.1. The idea behind the notation of the product integral comes from a Riemann type of construction using step-functions. If we approximate $\boldsymbol{M}(x)$ by a piecewise constant matrix function taking values $\boldsymbol{M}(x_i)$ on $[x_i, x_i + \Delta x_i)$ for $s = x_0 < x_1 < \cdots < x_N = t$ and where $\Delta x_i = x_{i+1} - x_i$, then by (5.2.6) the product integral over $[x_i, x_i + \Delta x_i)$ equals the matrix exponential

$$\mathrm{e}^{\boldsymbol{M}(x_i)\Delta x_i} = \boldsymbol{I} + \boldsymbol{M}(x_i)\Delta x_i + O(\Delta x_i^2).$$

By letting $\Delta x_i \to 0$ and using (5.2.3) we then arrive at the notation (5.2.2). △

A valuable formula for computing integrals involving product integrals is the so-called Van-Loan's formula for product integrals (see Lemma 2 in Bladt, Asmussen, and Steffensen, 2020), which states that

$$\prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{A}(u) & \boldsymbol{B}(u) \\ \boldsymbol{0} & \boldsymbol{C}(u) \end{pmatrix} \mathrm{d}u \right) \tag{5.2.9}$$

$$= \begin{pmatrix} \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(u)\,\mathrm{d}u) & \int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}(u)\,\mathrm{d}u) \boldsymbol{B}(x) \prod_x^t (\boldsymbol{I} + \boldsymbol{C}(u)\,\mathrm{d}u)\,\mathrm{d}x \\ \boldsymbol{0} & \prod_s^t (\boldsymbol{I} + \boldsymbol{C}(u)\,\mathrm{d}u) \end{pmatrix}.$$

This formula is valid for matrix functions $\boldsymbol{A}(x), \boldsymbol{B}(x)$ and $\boldsymbol{C}(x)$, which are piecewise continuous. The matrices $\boldsymbol{A}(x)$ and $\boldsymbol{C}(x)$ are square matrices of possibly different dimensions, so $\boldsymbol{B}(x)$ is not necessarily a square matrix.

Let

$$\boldsymbol{C}(s,t) = \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \otimes \boldsymbol{I},$$

where $\otimes$ denotes the Kronecker product. The Kronecker product between a $p_1 \times q_1$ matrix $\boldsymbol{A} = \{a_{ij}\}$ and a $p_2 \times q_2$ matrix $\boldsymbol{B} = \{b_{ij}\}$ is defined as the $p_1 p_2 \times q_1 q_2$ matrix

$$\boldsymbol{A} \otimes \boldsymbol{B} = \{a_{ij}\boldsymbol{B}\}_{i=1,\dots,p_1, j=1,\dots,q_1} = \{a_{ij} b_{k\ell}\}.$$

Using that $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$, we get

$$\begin{aligned} \frac{\partial}{\partial t}\boldsymbol{C}(s,t) &= \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{A}(t) \otimes \boldsymbol{I} \\ &= \left(\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \otimes \boldsymbol{I} \right) (\boldsymbol{A}(t) \otimes \boldsymbol{I}) \\ &= \boldsymbol{C}(s,t)\,(\boldsymbol{A}(t) \otimes \boldsymbol{I}), \end{aligned}$$

and we conclude that

$$\boldsymbol{C}(s,t) = \prod_s^t (\boldsymbol{I} + (\boldsymbol{A}(x) \otimes \boldsymbol{I})\,\mathrm{d}x). \tag{5.2.10}$$

A similar argument gives that

$$\boldsymbol{I} \otimes \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + (\boldsymbol{I} \otimes \boldsymbol{A}(x))\,\mathrm{d}x). \tag{5.2.11}$$

Finally, if $\boldsymbol{A}(t)$ and $\boldsymbol{B}(t)$ are Riemann integrable matrix functions of dimensions $q \times q$ and $p \times p$ respectively, then

$$\prod_s^t (\boldsymbol{I} + (\boldsymbol{A}(x) \oplus \boldsymbol{B}(x))\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \otimes \prod_s^t (\boldsymbol{I} + \boldsymbol{B}(x)\,\mathrm{d}x), \tag{5.2.12}$$

where $\oplus$ denotes the Kronecker sum, defined by $\boldsymbol{A}(t) \oplus \boldsymbol{B}(t) = \boldsymbol{A} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}(t)$, and where the first $\boldsymbol{I}$ has the dimension of $\boldsymbol{B}(t)$ and the second $\boldsymbol{I}$ has the dimension of $\boldsymbol{A}(t)$. To see this, we notice that $\boldsymbol{A}(t) \otimes \boldsymbol{I}$ and $\boldsymbol{I} \otimes \boldsymbol{B}(t)$ commute, so by (5.2.7) we get that

$$\begin{aligned}
\prod_s^t (\boldsymbol{I} + (\boldsymbol{A}(x) \oplus \boldsymbol{B}(x))\,\mathrm{d}x) &= \prod_s^t (\boldsymbol{I} + (\boldsymbol{A}(x) \otimes \boldsymbol{I})\,\mathrm{d}x) \prod_s^t (\boldsymbol{I} + (\boldsymbol{I} \otimes \boldsymbol{B}(x))\,\mathrm{d}x) \\
&= \left[\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \otimes \boldsymbol{I}\right] \left[\boldsymbol{I} \otimes \prod_s^t (\boldsymbol{I} + \boldsymbol{B}(x)\,\mathrm{d}x)\right] \\
&= \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \otimes \prod_s^t (\boldsymbol{I} + \boldsymbol{B}(x)\,\mathrm{d}x).
\end{aligned}$$

For further details on Kronecker products and sums, we refer to Graham (1981).

5.2.3 Phase-type distributions

Consider a (time-inhomogeneous) Markov jump process $\{Y(t)\}_{t\geq 0}$, where state $p+1$ is absorbing and $1, ..., p$ are transient. The intensity matrix $\boldsymbol{M}(x)$ for $\{Y(t)\}_{t\geq 0}$ is then on the form

$$\boldsymbol{M}(x) = \begin{pmatrix} \boldsymbol{T}(x) & \boldsymbol{t}(x) \\ \boldsymbol{0} & 0 \end{pmatrix}, \tag{5.2.13}$$

where $\boldsymbol{T}(x)$ is a $p \times p$ sub-intensity matrix consisting of transition rates between transient states, and $\boldsymbol{t}(x) = -\boldsymbol{T}(x)\boldsymbol{e}$ is a column vector of exit rates, i.e. rates for jumping to the absorbing state. Then by Van-Loan's formula (5.2.9), the transition matrix for $\{Y(t)\}_{t\geq 0}$ is given by

$$\begin{aligned}
\boldsymbol{P}(s,t) &= \prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{T}(u) & \boldsymbol{t}(u) \\ \boldsymbol{0} & 0 \end{pmatrix} \mathrm{d}u\right) \\
&= \begin{pmatrix} \prod_s^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u) & \boldsymbol{e} - \prod_s^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix}.
\end{aligned}$$

Hence $\prod_s^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)$ is the matrix which contains the transition probabilities between the transient states from times s to t.

We assume that $\mathbb{P}(Y(0) = p+1) = 0$, and define $\pi_i = \mathbb{P}(Y(0) = i)$. Hence $\boldsymbol{\pi} = (\pi_1, ..., \pi_p)$ satisfies that $\boldsymbol{\pi e} = \sum_i \pi_i = 1$, so that $\boldsymbol{\pi}$ is the initial distribution for $\{Y(t)\}_{t\geq 0}$ concentrated on the transient states only. Then

$$(\mathbb{P}(Y(t) = 1), \mathbb{P}(Y(t) = 2), ..., \mathbb{P}(Y(t) = p)) = \boldsymbol{\pi} \prod_0^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u) \tag{5.2.14}$$

is a row vector that contains the probabilities of the process being in the different transient states at time t.

Now let

$$\tau = \inf\{t > 0 : Y(t) = p+1\}$$

denote the time until absorption. Then from (5.2.14) we immediately get that

$$\mathbb{P}(\tau > t) = \boldsymbol{\pi} \prod_0^t (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{e} \tag{5.2.15}$$

since the right-hand side equals the probability of the process belonging to *any* of the transient states by time t, i.e., absorption has not yet occurred. Differentiating (5.2.15) and using (5.2.1) we see that τ has a density on the form

$$f_\tau(x) = \boldsymbol{\pi} \prod_0^x (\boldsymbol{I} + \boldsymbol{T}(u)\,\mathrm{d}u)\boldsymbol{t}(x). \tag{5.2.16}$$

Definition 5.2.2. *The distribution of τ is called an inhomogeneous phase-type distribution, and we write $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{T}(x))$, where the indexation of $\boldsymbol{T}(x)$ is over $x \geq 0$.*

We do not need to specify $\boldsymbol{t}(x)$ since it is implicitly given by $\boldsymbol{T}(x)$. Indeed, since row sums of intensity matrices (and hence of (5.2.13)) are zero, we have that $\boldsymbol{t}(x) = -\boldsymbol{T}(x)\boldsymbol{e}$. If $\boldsymbol{T}(x) \equiv \boldsymbol{T}$, then we simply write $\tau \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$. This corresponds to the underlying Markov jump process being time-homogeneous.

We also notice $\boldsymbol{T}(x) + \boldsymbol{\Delta}(\boldsymbol{t}(x))$ defines an intensity matrix (without the absorbing state).

The class of phase-type distributions (both PH and IPH) is dense (in the sense of weak convergence) in the class of distributions on the positive reals, implying that any distribution with support $\mathbb{R}_+$ may be approximated arbitrarily close by a phase-type distribution. This result is also of considerable practical importance since phase-type distributions can be fitted both to data and distributions using a maximum likelihood approach. For the time-homogenous case, PH, see Asmussen, Nerman, and Olsson (1996) while for IPH we refer to Albrecher, Bladt, and Yslas (2022).

5.3 Phase-type representations of bond prices

Consider the stochastic interest rate model of (5.1.1), and let $E = \{1, \dots, p\}$ denote the state-space of the Markov jump process $X = \{X(t)\}_{t\geq 0}$ with intensity matrix $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j\in E}$. Let $\boldsymbol{r}(t) = (r_1(t), \dots, r_p(t))'$ be the column vector which contains the interest rate functions.

The main result of this section is the following result.

Theorem 5.3.1. *For $i, j \in E$, let*

$$d_{ij}(s,t) = \mathrm{E}\left(\mathbb{1}_{(X(t)=j)} \exp\left(-\int_s^t r_{X(u)}(u)\,\mathrm{d}u \right) \middle| X(s) = i \right), \quad s \leq t.$$

Then the matrix $\boldsymbol{D}(s,t) = \{d_{ij}(s,t)\}_{i,j\in E}$ has the following representation

$$\boldsymbol{D}(s,t) = \prod_s^t (\boldsymbol{I} + [\boldsymbol{M}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u))]\,\mathrm{d}u). \tag{5.3.1}$$

Proof. Conditioning on the state of $s + \mathrm{d}s$, we get that

$$\begin{aligned} d_{ij}(s,t) &= (1 + \mu_{ii}(s)\,\mathrm{d}s) d_{ij}(s + \mathrm{d}s, t)(1 - r_i(s)\,\mathrm{d}s) \\ &\quad + \sum_{k\neq i} \mu_{ik}(s)\,\mathrm{d}s d_{kj}(s + \mathrm{d}s, t)(1 - r_i(s)\,\mathrm{d}s) \\ &= d_{ij}(s + \mathrm{d}s, t)(1 - r_i(s)\,\mathrm{d}s) + \mu_{ii}(s)\,\mathrm{d}s d_{ij}(s + \mathrm{d}s, t) \\ &\quad + \sum_{k\neq i} \mu_{ik}(s)\,\mathrm{d}s d_{kj}(s + \mathrm{d}s, t) \end{aligned}$$

so that

$$-\frac{\partial}{\partial s} d_{ij}(s,t) \quad = \quad -r_i(t) d_{ij}(s,t) + \sum_k \mu_{ik}(t) d_{kj}(s,t). \tag{5.3.2}$$

In matrix form, this amounts to

$$\frac{\partial}{\partial s}\boldsymbol{D}(s,t) = -\left(\boldsymbol{M}(s) - \boldsymbol{\Delta}(\boldsymbol{r}(s))\right)\boldsymbol{D}(s,t). \tag{5.3.3}$$

Noting that $\boldsymbol{D}(t,t) = \boldsymbol{P}(t,t) = \boldsymbol{I}$, we hence conclude that (5.3.1) holds. □

Remark 5.3.2. The quantities $d_{ij}(s,t)$ in Theorem 5.3.1 are introduced as $\boldsymbol{r}$-deflated transition probabilities in Buchardt, Furrer, and Møller (2020, Appendix 1), where the authors derive the differential equation (5.3.3). While they give a martingale-based proof, we provide a probabilistic sample path argument and give a product integral representation. △

Remark 5.3.3. Multiplying both sides of (5.3.3) with $\boldsymbol{e}$ from the right, we recover the differential equation for the state-wise discount factors obtained in Norberg (1995a, (4.4)). △

Assume that all $r_i(x)$ are bounded from below, and let

$$\rho = \max\left(0, -\min_{i\in E}\inf_{x\geq 0} r_i(x)\right).$$

Then $\rho = 0$ if all interest rates are non-negative, and otherwise $-\rho$ provides a lower bound for all of them. Then we have the following result.

Theorem 5.3.4. *The price of the zero-coupon bond* (5.1.2) *satisfies*

$$B(t,T) = \mathrm{E}^{\mathbb{Q}}\left(\exp\left(-\int_t^T r_{X(u)}(u)\,\mathrm{d}u\right)\middle| X(t)\right) = \boldsymbol{e}_{X(t)}'\boldsymbol{D}(t,T)\boldsymbol{e}. \qquad (5.3.4)$$

Conditional on $X(t) = i$,

$$T \to \mathrm{e}^{-\rho(T-t)}B(t,T)$$

is the survival function for an IPH distributed random variable, $\tau(t)$*, with initial distribution* $\boldsymbol{e}_i'$ *and intensity matrices* $\boldsymbol{M}(x+t) - \boldsymbol{\Delta}(\boldsymbol{r}(x+t)) - \rho\boldsymbol{I}$*,* $x \geq 0$*.*

In particular, if all interest rates are non-negative, then $\rho = 0$ *and the price itself,* $T \to B(t,T)$ *becomes the survival function.*

Proof. The formula (5.3.4) follows directly from the construction of the $\boldsymbol{D}(t,T)$ matrix by summing out over j in $d_{ij}(t,T)$, which corresponds to post-multiplying $\boldsymbol{D}(t,T)$ by $\boldsymbol{e}$. Next, we notice that

$$e^{-\rho(T-t)}\prod_t^T(\boldsymbol{I} + [\boldsymbol{M}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u))]\,\mathrm{d}u) = \prod_t^T(\boldsymbol{I} + [\boldsymbol{M}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u)) - \rho\boldsymbol{I}]\,\mathrm{d}u),$$

which follows from (5.2.8). The matrix $\boldsymbol{M}(x) - \boldsymbol{\Delta}(\boldsymbol{r}(x)) - \rho\boldsymbol{I}$ is a sub-intensity matrix, which together with the distribution for $X(t)$ defines a phase-type representation $(\boldsymbol{\pi}_t, \boldsymbol{M}(x+t) - \boldsymbol{\Delta}(\boldsymbol{r}(x+t) - \rho\boldsymbol{I})$, $x \geq 0$ (starting at time t). □

The forward rate $f(t,T)$ is defined by

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

Using Theorem 5.3.4, we may write

$$B(t,T) = \mathrm{e}^{\rho(T-t)}\bar{F}_{\tau(t)}(T),$$

where $\bar{F}_{\tau(t)}(T) = 1 - F_{\tau(t)}(T)$ denotes the survival function for

$$\tau(t) \sim \mathrm{IPH}(\boldsymbol{e}_{X(t)}', \boldsymbol{M}(x+t) - \boldsymbol{\Delta}(\boldsymbol{r}(x+t)) - \rho\boldsymbol{I}).$$

Then

$$-\frac{\partial}{\partial T}\log B(t,T) = -\rho + \frac{f_{\tau(t)}(T)}{1-F_{\tau(t)}(T)},$$

where $f_{\tau(t)}$ denotes the density function for $\tau(t)$. Hence we have proved the following result.

Corollary 5.3.5. *Conditional on $X(t) = i$, the forward rate $f(t,T)$ equals the hazard rate at T for the random variable $\tau(t) \sim \mathrm{IPH}(\boldsymbol{e}_i, \boldsymbol{M}(x+t) - \boldsymbol{\Delta}(\boldsymbol{r}(x+t)) - \rho\boldsymbol{I})$, less ρ, i.e.*

$$f(t,T) = \frac{f_{\tau(t)}(T)}{1-F_{\tau(t)}(T)} - \rho. \tag{5.3.5}$$

Another immediate consequence of Theorem 5.3.4 is the following.

Corollary 5.3.6. *Assume that all interest rates are non-negative. Then conditional on $X(t) = i$, the random variable $\tau(t) \sim \mathrm{IPH}(\boldsymbol{e}_i', \boldsymbol{M}(t+x) - \boldsymbol{\Delta}(\boldsymbol{r}(t+x)))$, $x \geq 0$ then has a c.d.f. given by*

$$F_{\tau(t)}(T) = 1 - B(t,T) = \mathrm{E}^{\mathbb{Q}}\left(\int_t^T r_{X(y)}(y)\mathrm{e}^{-\int_t^y r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}y \,\middle|\, X(t) = i\right).$$

Proof. This follows from Theorem 5.3.4with $\rho = 0$ and

$$f_{\tau(t)}(y) \quad = \quad -\frac{\partial}{\partial y}B(t,y) = \mathrm{E}^{\mathbb{Q}}\left(r_{X(y)}(y)\mathrm{e}^{-\int_t^y r_{X(u)}(u)\,\mathrm{d}u}\,\middle|\, X(t) = i\right).$$

Integrating the expression then yields the result. □

For the case where $t = 0$, the above results are reduced to the following.

Corollary 5.3.7. *Assume that all interest rates are non-negative. Let $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{M}(x) - \boldsymbol{\Delta}(\boldsymbol{r}(x)))$ and let $\boldsymbol{\pi} = (\pi_1, ..., \pi_p)'$ denote the (initial) distribution of $X(0)$. Then*

$$\mathbb{P}(\tau > T) \quad = \quad \mathrm{E}^{\mathbb{Q}}\left(\exp\left(-\int_0^T r_{X(u)}(u)\,\mathrm{d}u\right)\right) \tag{5.3.6}$$

$$F_{\tau(t)}(T) \quad = \quad = \mathrm{E}^{\mathbb{Q}}\left(\int_0^T r_{X(y)}(y)\mathrm{e}^{-\int_0^y r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}y\right) \tag{5.3.7}$$

$$f(0,T) \quad = \quad \frac{f_\tau(T)}{1-F_\tau(T)}. \tag{5.3.8}$$

Remark 5.3.8. The density $f_\tau(t)$ has the interpretation of being the expected present value of the current interest rate accumulated in a small time interval arround t, and $F_\tau(T)$ is the present value of the total accumulated interest rate during $[0,T]$. △

Example 5.3.9. Assume that all interest rates are non-negative. If $\{X(t)\}_{t\geq 0}$ is time-homogeneous and $\boldsymbol{r}(t) = \boldsymbol{r} = (r_1, ..., r_p)$, then we also have that

$$\begin{aligned}
\mathrm{E}^{\mathbb{Q}}\left(\int_0^T \mathrm{e}^{-\int_0^y r_{X(u)}\,\mathrm{d}u}\,\mathrm{d}y\right) &= \int_0^T \mathrm{E}^{\mathbb{Q}}\left(\mathrm{e}^{-\int_0^y r_{X(u)}\,\mathrm{d}u}\right)\mathrm{d}y \\
&= \int_0^T \mathbb{P}(\tau > y)\,\mathrm{d}y \\
&= \int_0^T \boldsymbol{\pi}\mathrm{e}^{(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))y}\boldsymbol{e}\,\mathrm{d}y \\
&= \boldsymbol{\pi}(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))^{-1}\mathrm{e}^{(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))T}\boldsymbol{e} - \boldsymbol{\pi}(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))^{-1}\boldsymbol{e} \\
&= \mu\left[1 - \tilde{\boldsymbol{\pi}}\mathrm{e}^{(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))T}\boldsymbol{e}\right] \\
&= \mu\mathbb{P}(\tilde{\tau} > T),
\end{aligned}$$

where $\mu = \boldsymbol{\pi}\left[-(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))\right]^{-1}\boldsymbol{e}$ is the expectation of τ,

$$\tilde{\boldsymbol{\pi}} = \frac{\boldsymbol{\pi}\left[-(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))\right]^{-1}}{\boldsymbol{\pi}\left[-(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))\right]^{-1}\boldsymbol{e}}$$

is the stationary distribution of a phase-type renewal process with inter-arrivals being $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))$, see Bladt and Nielsen (2017, Th. 5.3.4), and $\tilde{\tau} \sim \mathrm{PH}(\tilde{\boldsymbol{\pi}}, \boldsymbol{M} - \boldsymbol{\Delta}(\boldsymbol{r}))$. Hence the swap rate ρ can be expressed as

$$\begin{aligned}
\rho &= \frac{\mathrm{E}^{\mathbb{Q}}\left(\int_0^T r_{X(y)}\mathrm{e}^{-\int_0^y r_{X(u)}\,\mathrm{d}u}\,\mathrm{d}y\right)}{\mathrm{E}^{\mathbb{Q}}\left(\int_0^T \mathrm{e}^{-\int_0^y r_{X(u)}\,\mathrm{d}u}\,\mathrm{d}y\right)} = \frac{F_\tau(T)}{\mu\mathbb{P}(\tilde{\tau} > T)} \\
&= \frac{1 - \boldsymbol{\pi}\mathrm{e}^{(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))T}\boldsymbol{e}}{\boldsymbol{\pi}\left[-(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))\right]^{-1}\mathrm{e}^{(\boldsymbol{M}-\boldsymbol{\Delta}(\boldsymbol{r}))T}\boldsymbol{e}}.
\end{aligned}$$

△

5.4 Estimation

Time-homogeneous phase-type distributions or inhomogeneous phase-type distribution where the sub-intensity matrices are on the form

$$\boldsymbol{T}(x) = \lambda_\theta(x)\boldsymbol{T},$$

for some parametric function $\lambda_\theta(x)$, can be estimated in terms of an EM algorithm.

An observation from a phase-type distribution is hence considered to be the time until a Markov jump process is absorbed, where all transitions and sojourn times in the different states are unobserved. This makes the estimation an incomplete data problem, which an EM algorithm can solve. Essentially the unobserved sufficient statistics (number of jumps between states, total time in then different states) are replaced by their conditional expectations given data and used in the explicit formulas for the maximum likelihood estimators. This updates the parameters, and the procedure is repeated until convergence. Convergence is secured as the likelihood increases in each step. The limit may be a global or only a local maximum.

Repeated data (absorption times), of course result in the same conditional expectations given their data. This carries over to weighted data as well, and hence the EM algorithm may efficiently estimate data in histograms. In particular, we may estimate to theoretical distributions by treating their discretized density as a histogram. This provides the link to fitting the intensity matrix of $\{X(t)\}_{t\geq 0}$ in (5.1.1) through bond prices, (5.1.2) or (5.1.3), either in terms of observed data or to a theoretical model.

Indeed, consider bond prices $B(0,T_i)$ available at different maturities $T_1,...,T_n$. Then according to Theorem 5.3.4 we have that

$$B(0,T_i) = \boldsymbol{\pi}\boldsymbol{D}(0,T_i)\boldsymbol{e} = \mathrm{e}^{\rho T_i}\mathbb{P}(\tau > T_i), \quad i = 1,2,...,n,$$

for some $\rho > 0$ and where $\tau \sim \mathrm{IPH}(\boldsymbol{\pi}, \boldsymbol{M}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u)) - \rho\boldsymbol{I})$. Then ρ must satisfy that

$$\mathrm{e}^{-\rho T_i}B(0,T_i) \leq 1, \quad i = 1,2,...,n.$$

This can be achieved by choosing

$$\rho = \max_{i\in\{1,\dots,n\}} \left(\frac{\log B(0,T_i)}{T_i}\right).$$

In the life insurance context in Denmark, by regulation the bond prices (discounting factors) must be computed from discrete forward rates, $f_d(0,T_i)$, published by the Danish Financial Supervisory Authority. Thus

$$B(0,T_i) = (1 + f_d(0,T_i))^{-T_i}$$

from which

$$\frac{\log B(0,T_i)}{T_i} = -\log(1 + f_d(0,T_i)).$$

Hence

$$\rho = \max_i\left(-\log(1 + f_d(0,T_i))\right) = -\min_i \log(1 + f_d(0,T_i)). \tag{5.4.1}$$

Hence calibrating to data $B(0,T_i)$, $i = 1,...,n$ can be done by fitting PH or IPH distributions to $\mathrm{e}^{-\rho T_i}B(0,T_i)$ using an EM algorithm. The possible interest rates can either be picked by the EM algorithm (referred to as unrestricted interest rates), or we can fix the possible rates to values (or functions) of our choice (restricted interest rates).

In the former case, we obtain a maximum likelihood estimate $(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{T}}(x))$ for the parameters. The estimate for $\boldsymbol{M}(x)$ is then readily obtained from

$$\hat{\boldsymbol{M}}(x) = \hat{\boldsymbol{T}}(x) + \boldsymbol{\Delta}(\boldsymbol{t}(x)).$$

To find the induced interest rates, we also have from Theorem 5.3.4 that

$$\hat{\boldsymbol{T}}(x) = \hat{\boldsymbol{M}}(x) - \boldsymbol{\Delta}(\boldsymbol{r}(x)) - \rho \boldsymbol{I}$$

so we conclude that the estimated exit rates $\boldsymbol{t}(x)$ must satisfy

$$\boldsymbol{t}(x) = \boldsymbol{r}(x) + \rho \boldsymbol{e},$$

where $\boldsymbol{e}$ is the vector of ones. Hence the induced interest rates are given by

$$\boldsymbol{r}(x) = \boldsymbol{t}(x) - \rho \boldsymbol{e}.$$

Neither the transition rates nor the interest rates are unique, but the resulting discount factor (bond price) is invariant under different representations, which is all that matters regarding reserving in the insurance context.

If, in turn, we decide to choose the possible range of interest rates $r_i(x)$ ourselves, then the EM-algorithm is modified not to update the exit rates. This modification is easily dealt with by simply removing updates of the latter in the original EM algorithm of Asmussen, Nerman, and Olsson (1996) or Albrecher, Bladt, and Yslas (2022). See Appendix 5.A for details. In this case, the exit rates will be fixed at

$$\boldsymbol{t}(x) = \boldsymbol{r}(x) + \rho \boldsymbol{e}$$

so

$$\hat{\boldsymbol{M}}(x) = \hat{\boldsymbol{T}}(x) + \boldsymbol{\Delta}(\boldsymbol{r}(x)) + \rho \boldsymbol{I}.$$

While the parametrization of the transition rates may not be unique, the interest rates remain fixed.

We now present two examples of fitting to real data and one example to a theoretical model. The estimation is computed using the R-package *matrixdist*.

Example 5.4.1 (Fitting to observed bond prices with restricted interest rates)**.** Bond prices, $B(0, T)$ as of 31/12/2003 (time zero) with maturities $T = 1, 2, ..., 30$ years are available from the Danish Financial Supervisory Authority and given by 0.9755051, 0.9434934, 0.9059545, 0.8679149, 0.8251354, 0.7857250, 0.7472528, 0.7075066, 0.6679984, 0.6286035, 0.5951316, 0.5625969, 0.5310441, 0.5005108, 0.4710280, 0.4448469, 0.4197550, 0.3958013, 0.3728296, 0.3508858, 0.3319907, 0.3140894, 0.2970098, 0.2808430, 0.2654229, 0.2508400, 0.2369349, 0.2237965, 0.2112725, 0.1994495, respectively.

This corresponds to an empirical survival distribution to which we can then fit phase-type distributions of different dimensions. Regarding the discretization, we let $0.5 + i$, $i = 0, ..., 29$ denote the data points with probability mass $B(i) - B(i+1)$, where $B(0) = 1$, and a right censored data point at 30 with probability mass $B(30) = 0.1994495$. Since all observed bond prices are less than one, we have $\rho = 0$, corresponding to an environment with non-negative interest rates.

We used $p = 2, 3, 4, 5, 10$ and 15 phases, with state-wise interest rates being $r_i^p = i/(10p)$, $i = 1, ..., p$ for the different dimensions p. Underlying this choice is the assumption that the interest rates fluctuate between 1% and 10%, and the r_i's are obtained as the points that divide the interval $[0, 0.1]$ into p, including the right endpoint. The vectors $\boldsymbol{r}^p = (r_1^p, ..., r_p^p)'$ will serve as exit rate vectors of the phase-type distributions to be fitted.

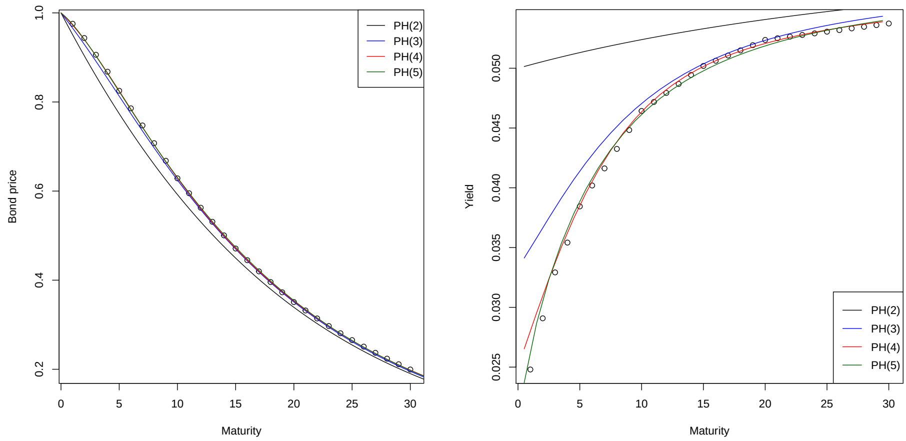

Figure 5.1: *Phase-type fits to Zero-coupon bond prices (left) and corresponding yield curves (right) for dimension* $p = 2, 3, 4, 5$.

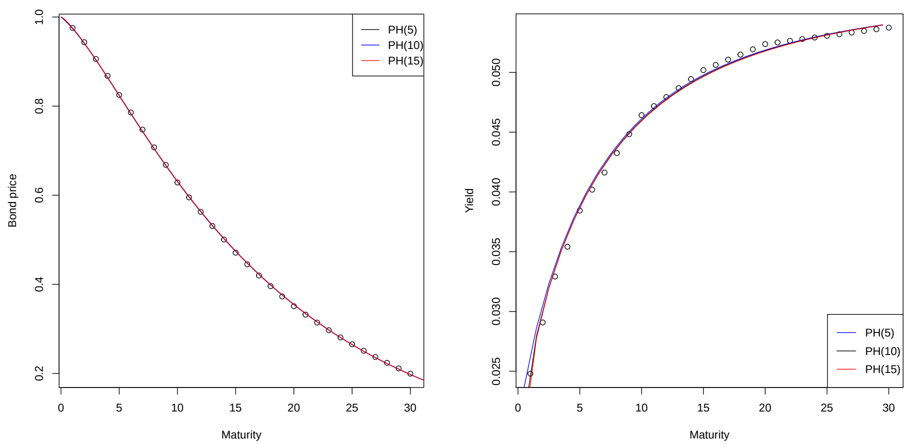

Figure 5.2: *Phase-type fits to Zero-coupon bond prices (left) and corresponding yield curves (right) for dimension* $p = 5, 10, 15$.

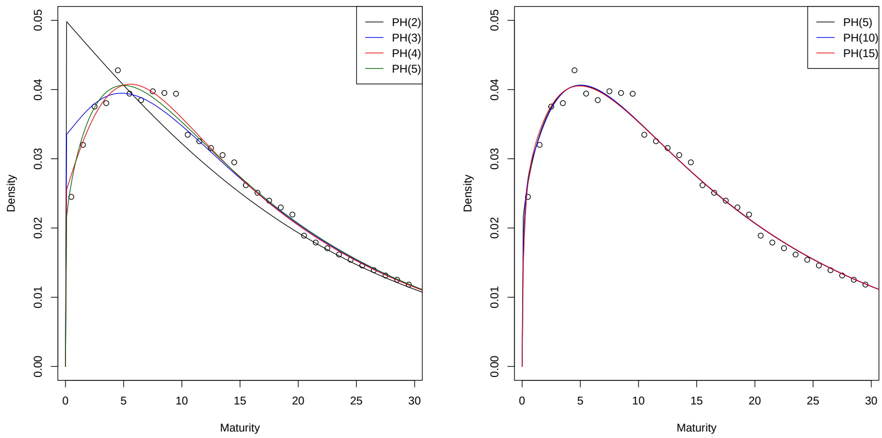

Figure 5.3: *Fitted phase-type densities vs. weighted data for* $p = 2, 3, 4, 5$ *(left) and* $p = 5, 10, 15$ *(right).*

In Figure 5.1 (left), we have plotted the phase-type fits to the empirical survival curve for dimensions $p = 2, 3, 4, 5$. At dimension 3, we obtain a decent fit and excellent fits for dimensions 4 and 5. The likelihood values for 4 cases are -3.178171, -3.16838,-3.166633, and -3.166182. Further experimentation with dimensions 10 and 15 resulted in likelihoods of -3.165002 and -3.164654, respectively. However, the plots of bond prices and yields are indistinguishable from the plots corresponding to dimension 5, see Figure 5.2. We can also assess the quality of the fits by plotting the estimated density function vs. the weighted data, shown in Figure 5.3. Again the plots for dimensions $p = 5, 10, 15$ are almost indistinguishable. Therefore, we conclude that dimension 4 or 5 will suffice to *approximate* the bond prices.

The estimates of the sub-intensity matrix $\boldsymbol{M} - \boldsymbol{\Delta}(\boldsymbol{r})$ (under a risk neutral measure $\mathbb{Q}$) for dimensions $p = 3, 4, 5$ are given by

$$\begin{pmatrix} -0.13 & 0.1 & 0 \\ 0 & -0.41 & 0.34 \\ 0.14 & 0 & -0.24 \end{pmatrix}, \begin{pmatrix} -0.25 & 0.22 & 0.01 & 0 \\ 0.14 & -1.11 & 0.75 & 0.18 \\ 0.06 & 0.29 & -0.63 & 0.2 \\ 0.09 & 0.22 & 0.65 & -1.05 \end{pmatrix},$$

and

$$\begin{pmatrix} -0.26 & 0.02 & 0.06 & 0.07 & 0.08 \\ 0.07 & -1.68 & 0.69 & 0.23 & 0.65 \\ 0.19 & 0.32 & -1.99 & 0.93 & 0.48 \\ 0.04 & 0.35 & 0.27 & -1.2 & 0.46 \\ 0.07 & 0.82 & 0.07 & 0.8 & -1.85 \end{pmatrix}.$$

To fit the bond prices, the initial distributions of Markov processes were all on the form $(1, 0, ..., 0)$ of appropriate dimension, i.e., initiation in state 1. ◦

Example 5.4.2 (Fitting to 2019 bond prices with unrestricted interest rates)**.** To illustrate the applicability of our methods also in the case of a negative interest rate environment, we can instead fit to bond prices as of 31/12/2019 from the Danish Financial Supervisory Authority; this dataset consists of maturities of $T = 1, 2, ..., 120$ years. In this case, we let the EM algorithm choose the necessary positive and negative interest rates.

The first five years have bond prices above one and given by 1.00231736, 1.00403337, 1.00445679, 1.00382807, and 1.00197787, which reflects the (slightly) negative interest rate environment at the time. From (5.4.1), we get $\rho = 0.002314677$ as the exponential factor to down-scale prices to below one.

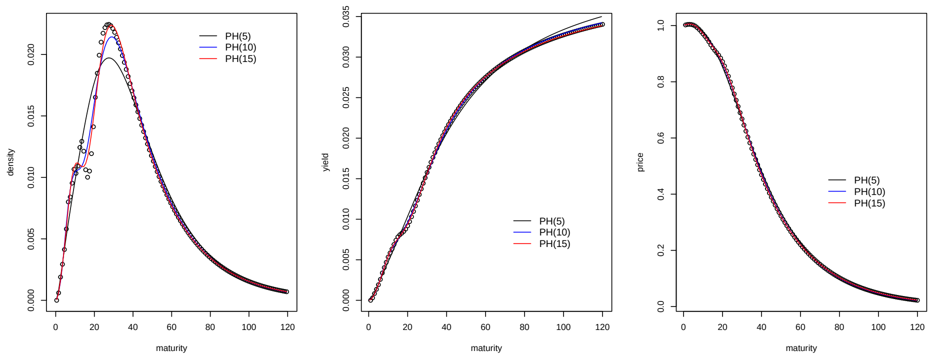

Figure 5.4: *Fitted phase-type densities (left), corresponding yield curves (middle) and bond prices (right) for dimensions $p = 5{,}10{,}15$ based on bond price data as of 31/12/2019.*

In Figure 5.4, we show the phase-type fits to the bond prices. We have used the subclass of time-homogeneous Coxian distributions, where initiation is always in state 1, and the only possible transitions are from a state, i say, to the following, $i + 1$, or to exit to the absorbing state.

If the primary purpose is using the fits as a discounting factor in a life insurance model, then probably all fits could be used (right plot). If the yield curve fitting is the concern, then only dimensions 10 and 15 seem to catch the appropriate curvature. Regarding the probability density of the phase-type, the 15-dimensional fit is the best.

To exemplify, we consider the ten dimensional fit. The fitted intensity matrix,

$\hat{\boldsymbol{M}}$, for $\{X(u)\}_{u\geq 0}$, is given by

$$\begin{pmatrix} -0.5212 & 0.5212 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.5212 & 0.5212 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & -0.5185 & 0.5185 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & -0.5161 & 0.5161 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.5152 & 0.5152 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.4664 & 0.4664 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3099 & 0.3099 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3099 & 0.3099 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3099 & 0.3099 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix}$$

The matrix contains six different parameter values. The matrix structure is carried over from the phase-type fit to the (discounted) bond prices. The blocks with the same parameters correspond to Erlang blocks, i.e. convolution of exponential distributions with the same parameter.

The induced (estimated) interest rates (in %) are, respectively,

$$-\rho, -\rho, 0.03468739, 0.28218594, -\rho, 4.64627655, -\rho, -\rho, -\rho, 3.86252219.$$

These should also be counted as parameters. ○

Example 5.4.3 (Fitting to a two-factor Vasicek model)**.** In this example we consider the two-factor Vasicek short rate model G2++ (see Brigo and Mercurio, 2006) with an initial negative interest rate.

Here the bond prices as of time zero are given by

$$B(0,T) = \exp\left\{-\psi(T) + \frac{1}{2}V^2(0,T)\right\},$$

where

$$\begin{aligned} V^2(0,T) = &\sum_{i=1}^{2} \frac{\sigma_i^2}{k_i^2}\left(T - t - B_{k_i}(0,T) - \frac{k_i}{2}B_{k_i}^2(0,T)\right) \\ &+ \frac{2\sigma_1\sigma_2\sigma_{12}}{k_1k_2}\left(T - t - B_{k_1}(0,T) - B_{k_2}(0,T) + B_{k_1+k_2}(0,T)\right), \end{aligned}$$

$$B_k(0,T) = \frac{1 - \mathrm{e}^{-k(T-t)}}{k} \quad \text{and} \quad \psi(T) = \frac{(\theta - r_0)(1 + \mathrm{e}^{-k_1T}) + k_1\theta T}{k_1}.$$

We chose the same parameters as in Diez and Korn (2020), Fig. 3, apart from the initial interest rate r_0, which was set to -1%. Hence the parameters are

$$\begin{aligned} &r_0 = -0.01,\ k_1 = 0.401,\ k_2 = 0.178,\ \sigma_1 = 0.0378,\ \sigma_2 = 0.0372, \\ &\theta = 0.01297,\ \sigma_{12} = -0.996. \end{aligned}$$

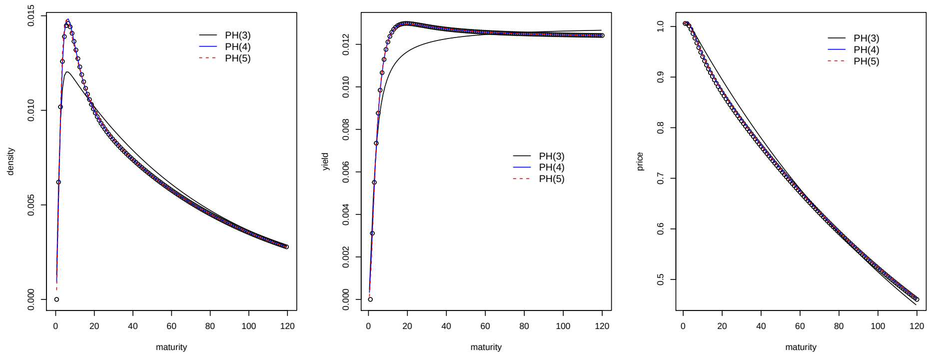

Figure 5.5: *Fitted phase-type densities (left) and corresponding yield curves (middle) and bond prices (right) for dimensions $p = 3,4,5$ based on bond prices from the two-factor Vasicek G2++ model.*

We fitted 3,4 and 5 dimensional time-homogeneous phase-type distributions with a Coxian structure to the discounted bond prices $\mathrm{e}^{-\rho T}B(0,T)$. Here $\rho = 0.005955398$ and the intensity matrix for $\boldsymbol{M}$ based on 4 phases is given by

$$\hat{\boldsymbol{M}} = \begin{pmatrix} -0.17 & 0.17 & 0.00 & 0.00 \\ 0.00 & -0.66 & 0.66 & 0.00 \\ 0.00 & 0.00 & -0.61 & 0.61 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{pmatrix}$$

with corresponding interest rates $-\rho, -\rho, 0.0782987520.006307674$, while for 5 phases, we get

$$\hat{\boldsymbol{M}} = \begin{pmatrix} -0.65 & 0.65 & 0.00 & 0.00 & 0.00 \\ 0.00 & -1.79 & 1.79 & 0.00 & 0.00 \\ 0.00 & 0.00 & -1.89 & 1.89 & 0.00 \\ 0.00 & 0.00 & 0.00 & -0.12 & 0.12 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{pmatrix}.$$

The corresponding (estimated) interest rates are

$$-\rho,\ -\rho,\ -\rho,\ 0.012793967,\ 0.006280658.$$

A total of six parameters specify the four-dimensional model, while seven parameters determine the five-dimensional. ◦

5.5 Applications to life insurance

In this section, we incorporate the stochastic interest rate model of the previous sections to life insurance valuations. We consider the model introduced by Norberg

(1995b,a) and extend their results on reserves and higher order moments to so-called partial reserves and higher order moments, that is, corresponding results on events of the terminal state. Partial reserves and moments play important roles when dealing with so-called retrospective reserves in single states (cf. Subsection 5.E in Norberg, 1991), which, however, is outside the scope of the present paper. We provide this extension following the matrix approach of Bladt, Asmussen, and Steffensen (2020) so that these types of results are extended to allow for stochastic interest rates on the form (5.1.1). The extensions of the results of these papers are pointed out in a series of remarks throughout the section.

5.5.1 A Life insurance model with stochastic interest rates

Let $X = \{X(t)\}_{t\geq 0}$ be a time-inhomogeneous Markov jump process with a finite state-space E and intensity matrix $\mathbf{\Lambda}(t) = \{\lambda_{ij}(t)\}_{i,j\in E}$. Then we define a payment process $\{B(t)\}_{t\geq 0}$ by

$$\mathrm{d}B(t) = \sum_{i\in E}\left(\mathbb{1}_{(X(t-)=i)}b_i(t)\,\mathrm{d}t + \sum_{j\in E} b_{ij}(t)\,\mathrm{d}N_{ij}(t)\right), \tag{5.5.1}$$

where $b_i(t)$ are continuous payment rates (negative if premiums) and $b_{ij}(t)$ lump sum payments, which occur according to the counting measure $N_{ij}(t)$. The intensity matrix is decomposed into

$$\mathbf{\Lambda}(t) = \mathbf{\Lambda}^0(t) + \mathbf{\Lambda}^1(t), \tag{5.5.2}$$

where $\mathbf{\Lambda}^1(t)$ is a non-negative matrix and, consequently, $\mathbf{\Lambda}^0(t)$ a sub-intensity matrix, i.e. row sums are non-positive. The counting process is linked to the transitions of X in the following way. Upon transition from i to j, $i \neq j$, in X at time t, a lump sum payment of $b_{ij}(t)$ will be triggered with probability

$$\frac{\lambda_{ij}^1(t)}{\lambda_{ij}^0(t) + \lambda_{ij}^1(t)}. \tag{5.5.3}$$

If $i = j$, then $N_{ii}(t)$ denotes an inhomogeneous Poisson process with intensity $\lambda_{ii}(t)$, and a lump sum during a sojourn in state i will then be triggered in $[t, t + \mathrm{d}t)$ with probability $\lambda_{ii}^1(t)\,\mathrm{d}t$.

Finally, we assume that the spot interest rates in state i follow a deterministic function $r_i(t)$. Hence the interest rates follow the model (5.1.1).

Remark 5.5.1. The classic Markov chain life insurance setting of, e.g., Hoem (1969a) and Norberg (1991), is the recovered if $r_i(t) \equiv r(t)$, $b_{ii}(t) = 0$ and if the probabilities (5.5.3) are either zero or one. Extending the classic setting to allow for different interest rates in the different states was considered in Norberg (1995b,a), where Thiele type of differential equations for the reserves and higher order moments were derived. $\triangle$

For the purpose of computing reserves and higher order moments, Bladt, Asmussen, and Steffensen (2020, (3.8)-(3.11)), we let $\boldsymbol{b}(t) = (b_i(t))_{i\in E}$ denote the vector containing the continuous rates, and define matrices

$$
\begin{aligned}
\boldsymbol{B}(t) &= \{b_{ij}(t)\}_{i,j\in E}\,, \\
\boldsymbol{R}(t) &= \boldsymbol{\Lambda}^1(t) \bullet \boldsymbol{B}(t) + \boldsymbol{\Delta}(\boldsymbol{b}(t)), \\
\boldsymbol{C}^{(k)}(t) &= \boldsymbol{\Lambda}^1(t) \bullet \boldsymbol{B}^{\bullet k}(t), \quad k \geq 2,
\end{aligned}
$$

where $\boldsymbol{\Delta}(\boldsymbol{b}(t))$ denotes the diagonal matrix with $\boldsymbol{b}(t)$ as diagonal. The operator $\bullet$ denotes Schur (entrywise) matrix product, defined by $\boldsymbol{A} \bullet \boldsymbol{B} = \{a_{ij}b_{ij}\}$ for matrices $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{B} = \{b_{ij}\}$.

Hence $\boldsymbol{B}(t)$ is the matrix containing the lump payments at transitions and at Poisson arrivals during sojourns, $\boldsymbol{R}(t)$ is the matrix whose ij'th element is the expected reward accumulated during $[t, t + \mathrm{d}t)$ upon transition from i to j, or during a sojourn in state i if $i = j$. The $\boldsymbol{C}^{(k)}(t)$ matrix is more technical to be used when dealing with higher order moments.

Finally, we let

$$
\boldsymbol{r}(t) = (r_i(t))_{i\in E}.
$$

denote the vector of interest rates.

Now assume that the interest rate process is modeled and fitted using bond prices like in Section 5.4. Accordingly there is a Markov jump process $X_r = \{X_r(t)\}_{t\geq 0}$ with state-space $E_r = \{1, 2, ..., p\}$ and intensity matrix $\boldsymbol{\Lambda}_r(t) = \{\lambda^r_{ij}(t)\}_{t\geq 0}$, say, such that the corresponding bond prices $B(t,T)$ are given as in Theorem 5.3.4. Similarly, we let $X_b = \{X_b(t)\}_{t\geq 0}$ denote the Markov jump process governing the transition between the biometric states with the state-space $E_b = \{1, 2, ..., q\}$ and intensity matrix $\boldsymbol{\Lambda}_b(t) = \{\lambda^b_{ij}(t)\}_{t\geq 0}$. Hence the Markov jump process appearing in (5.5.1) can be written on the form

$$
X(t) = (X_b(t), X_r(t)) \tag{5.5.4}
$$

with state-space $E = E_b \times E_r$.

Hence we need to decide upon an ordering of E, which will be lexicographical. This means the for elements $(i, \tilde{i}), (j, \tilde{j}) \in E$,

$$
(i, \tilde{i}) < (j, \tilde{j}) \iff (i-1)p + \tilde{i} < (j-1)p + \tilde{j}.
$$

In other words, each biometric state i consists of sub-states $(i, 1), ..., (i, q)$ depending on the state of the underlying Markov process X_r, see Figure 5.6. The processes X_b and X_r may or may not be independent, and the payment processes (5.5.1) likewise may or may not be independent of X_r. In the independent case the processes X_b

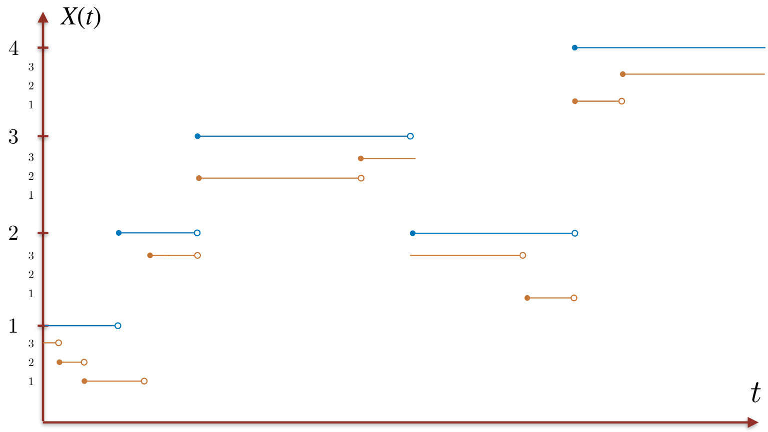

Figure 5.6: *Lexicographical ordering: for each biometric state (blue), several sub-states (orange) define the underlying interest rate level.*

and X_r are defined on each their state-space, and the common state-space will be the product set of the two. If the processes are sharing states, with the possibility of having simultaneous jumps, then we obtain dependency of the processes. Such a case could, e.g. be a rise in the interest rate causing an increased intensity of jumping to surrender or free-policy states (see, e.g., Buchardt, 2014).

In the following example, we consider the simplifications in the representations when assuming independence.

Example 5.5.2 (Independence)**.** If the transition rates of X satisfy, for all $i, j \in E_b$, $j \neq i$, and $\tilde{i}, \tilde{j} \in E_r$, $\tilde{j} \neq \tilde{i}$,

$$\lambda_{(i,\tilde{i}),(j,\tilde{i})}(t) = \lambda_{(i,\tilde{j}),(j,\tilde{j})}(t) = \lambda^b_{ij}(t),$$

$$\lambda_{(i,\tilde{i}),(i,\tilde{j})}(t) = \lambda_{(j,\tilde{i}),(j,\tilde{j})}(t) =: \lambda^r_{\tilde{i}\tilde{j}}(t),$$

we have that X_b and X_r are independent. Using the lexicographical ordering, we can, in this case, obtain compact matrix representations in terms of the two processes as follows. The transition intensity matrix of X is now of the form

$$\boldsymbol{\Lambda}(t) = \boldsymbol{\Lambda}_b(t) \oplus \boldsymbol{\Lambda}_r(t) = \boldsymbol{\Lambda}_b(t) \otimes \boldsymbol{I}_p + \boldsymbol{I}_q \otimes \boldsymbol{\Lambda}_r(t),$$

where $\oplus$ denotes the Kronecker sum, and where $\boldsymbol{I}_n$ denotes the identity matrix of dimension $n \times n$. We recall that the Kronecker product, $\otimes$, is defined by $\boldsymbol{A} \otimes \boldsymbol{B} = \{a_{ij}\boldsymbol{B}\}$, where $\boldsymbol{A} = \{a_{ij}\}$.

The interest rate vector satisfies

$$\boldsymbol{r}(t) = \boldsymbol{e} \otimes (r_1(t), ..., r_p(t)),$$

where $\boldsymbol{e} = (1, 1, ..., 1)'$.

If we further assume that the payment process (5.5.1) is independent of X_r, i.e. such that the payment functions satisfy, for all $i, j \in E_b$ and $\tilde{i}, \tilde{j} \in E_r$,

$$b_{(i,\tilde{i})}(t) = b_{(i,\tilde{j})}(t) =: b_i^b(t),$$

$$b_{(i,\tilde{i}),(j,\tilde{i})}(t) = b_{(i,\tilde{j}),(j,\tilde{j})}(t) =: b_{ij}^b(t),$$

$$b_{(i,\tilde{i}),(i,\tilde{j})}(t) = b_{(j,\tilde{i}),(j,\tilde{j})}(t) = 0,$$

we have that the payment matrices are on the form

$$\boldsymbol{B}(t) = \boldsymbol{B}^b(t) \otimes \boldsymbol{I}$$

$$\boldsymbol{b}(t) = \boldsymbol{b}^b(t) \otimes \boldsymbol{e}$$

where

$$\boldsymbol{b}^b(t) = \left(b_1^b(t), \dots, b_q^b(t)\right)' \qquad \text{and} \qquad \boldsymbol{B}(t) = \left\{b_{ij}^b(t)\right\}_{i,j \in E_b}.$$

Similarly, we may directly decompose $\boldsymbol{\Lambda}_b$:

$$\boldsymbol{\Lambda}_b(t) = \boldsymbol{\Lambda}_b^1(t) \oplus \boldsymbol{\Lambda}_b^0(t)$$

such that the decomposition (5.5.2) reads

$$\boldsymbol{\Lambda}^1(t) = \boldsymbol{\Lambda}_b^1(t) \otimes \boldsymbol{I}_q, \quad \text{and} \quad \boldsymbol{\Lambda}^0(t) = \boldsymbol{\Lambda}_b^0 \otimes \boldsymbol{I}_p + \boldsymbol{I}_q \otimes \boldsymbol{\Lambda}_r(t) = \boldsymbol{\Lambda}_b^0(t) \oplus \boldsymbol{\Lambda}_r(t),$$

The conceptual difference in the decomposition of $\boldsymbol{\Lambda}^1$ and $\boldsymbol{\Lambda}^0$ lies in the absence of lump sum payments upon transition between interest levels. ○

5.5.2 Reserves

We now consider the valuation of the payment process B. Introduce the matrix of partial state-wise prospective reserves,

$$\boldsymbol{V}(s,t) = \{V_{ij}(s,t)\}_{i,j \in E},$$

$$V_{ij}(s,t) = \mathrm{E}\left(\mathbb{1}_{(X(t)=j)} \int_s^t \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}B(x) \,\middle|\, X(s) = i\right).$$

Due to the stochastic interest rates, this is an extension of Bladt, Asmussen, and Steffensen (2020). With $\boldsymbol{D}(s,t)$, introduced in (5.3.1), modified to the setup of this section as

$$\boldsymbol{D}(s,t) = \prod_s^t (\boldsymbol{I} + [\boldsymbol{\Lambda}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u))]\,\mathrm{d}u),$$

we have the following result.

Theorem 5.5.3. *The matrix of partial state-wise prospective reserves* $\boldsymbol{V}(s,t)$ *has the following integral representation:*

$$\boldsymbol{V}(s,t) = \int_s^t \boldsymbol{D}(s,x)\boldsymbol{R}(x)\boldsymbol{P}(x,t)\,\mathrm{d}x. \tag{5.5.5}$$

Proof. See Appendix 5.B. □

The actual computation of the reserves can be effectively executed using the following Van-Loan type of formula, which avoids integration.

Corollary 5.5.4. $\boldsymbol{V}(s,t)$ *can be extracted from the relation*

$$\prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{\Lambda}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u)) & \boldsymbol{R}(u) \\ \boldsymbol{0} & \boldsymbol{\Lambda}(u) \end{pmatrix} \mathrm{d}u \right) = \begin{pmatrix} \boldsymbol{D}(s,t) & \boldsymbol{V}(s,t) \\ \boldsymbol{0} & \boldsymbol{P}(s,t) \end{pmatrix}.$$

Finally, we state and prove Thiele's differential equations for partial reserves with stochastic interest rates.

Theorem 5.5.5 (Thiele)**.**

$$\frac{\partial}{\partial s}\boldsymbol{V}(s,t) = -\left[\boldsymbol{\Lambda}(s) - \boldsymbol{\Delta}(\boldsymbol{r}(s))\right]\boldsymbol{V}(s,t) - \boldsymbol{R}(s)\boldsymbol{P}(s,t),$$

where $\boldsymbol{V}(t,t) = \boldsymbol{0}$. *For the conventional state-wise prospective reserves,* $\boldsymbol{V}^{Th}(t) = \boldsymbol{V}(t,T)\boldsymbol{e}$, *this has the form*

$$\frac{\partial}{\partial t}\boldsymbol{V}^{Th}(t) = \boldsymbol{\Delta}(\boldsymbol{r}(t))\boldsymbol{V}^{Th}(t) - \boldsymbol{\Lambda}(t)\boldsymbol{V}^{Th}(t) - \boldsymbol{R}(t)\boldsymbol{e},$$

where $\boldsymbol{V}^{Th}(T) = \boldsymbol{0}$.

Proof. See Appendix 5.B. □

Remark 5.5.6. Writing out the elements of the differential equation for $\boldsymbol{V}^{Th}$, we get for $i \in E$,

$$\frac{\partial}{\partial t}V_i^{Th}(t) = r_i(t)V_i^{Th}(t) - b_i(t) - \sum_{j\in E}\lambda_{ij}(t)\big(b_{ij}(t) + V_j^{Th}(t) - V_i^{Th}(t)\big),$$

$$V_i^{Th}(T) = 0,$$

which is the differential equation obtained in Norberg (1995b, (3.2)) in the case of a first-order moment. △

5.5.3 Higher order moments

Consider the matrix of partial state-wise higher order moments of future payments, given by, for $k \in \mathbb{N}$ (see Bladt, Asmussen, and Steffensen (2020, (3.6)-(3.7))),

$$\boldsymbol{V}^{(k)}(t,T) = \left\{V_{ij}^{(k)}(t,T)\right\}_{i,j\in E},$$

$$V_{ij}^{(k)}(t,T) = \mathrm{E}\left(\mathbb{1}_{(X(T)=j)}\left(\int_t^T \mathrm{e}^{-\int_t^x r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}B(x)\right)^k \,\middle|\, X(t)=i\right),$$

and introduce what we shall term the *reduced* partial state-wise higher order moments:

$$\boldsymbol{V}_r^{(k)}(t,T) = \frac{\boldsymbol{V}^{(k)}(t,T)}{k!}.$$

Since all payment functions and transition rates are deterministic, results for these higher-order moments are now straightforward to obtain by using the undiscounted result,

$$\begin{aligned}\boldsymbol{m}_r^{(k)}(t,T) &= \int_t^T \boldsymbol{P}(t,x)\boldsymbol{R}(x)\boldsymbol{m}_r^{(k-1)}(x,T)\ \mathrm{d}x \\ &\quad + \sum_{m=2}^{k}\int_t^T \boldsymbol{P}(t,x)\boldsymbol{C}_r^{(m)}(x)\boldsymbol{m}_r^{(k-m)}(x,T)\ \mathrm{d}x,\end{aligned}$$

where $\boldsymbol{m}_r^{(k)}(t,T)$, $k\in\mathbb{N}$, contains the partial state-wise k'th moment, normalised by $k!$, of the undiscounted future payments (see Bladt, Asmussen, and Steffensen (2020, (7.4))), i.e. $\boldsymbol{V}_r^{(k)}(s,T)$ with no interest rate. Indeed, rates $b_i(t)$ and lump sums $b_{ij}(t)$ must be replaced by the discounted versions with discounting factor, $\exp(-\int_s^t r_{X(u)}(u)\,\mathrm{d}u)$ (for fixed $s\le t$). Powers of lumps sums like $b_{ij}(t)^m$, $m\in\mathbb{N}$, are discounted by $\exp(-m\int_s^t r_{X(u)}(u)\,\mathrm{d}u)$. Denoting

$$\boldsymbol{D}^{(m)}(s,t) = \prod_s^t(\boldsymbol{I} + [\boldsymbol{\Lambda}(u) - m\boldsymbol{\Delta}(\boldsymbol{r}(u))]\ \mathrm{d}u), \quad m\in\mathbb{N},$$

we then obtain the following version of Hattendorff's theorem for partial reserves with stochastic interest rate.

Theorem 5.5.7. *The matrix of reduced partial state-wise higher order moments satisfies the integral equation, for $k\in\mathbb{N}_0$,*

$$\begin{aligned}\boldsymbol{V}_r^{(k)}(t,T) &= \int_t^T \boldsymbol{D}^{(k)}(t,x)\boldsymbol{R}(x)\boldsymbol{V}_r^{(k-1)}(x,T)\,\mathrm{d}x \\ &\quad + \sum_{m=2}^{k}\int_t^T \boldsymbol{D}^{(k)}(t,x)\boldsymbol{C}_r^{(m)}(x)\boldsymbol{V}_r^{(k-m)}(x,T)\,\mathrm{d}x.\end{aligned}$$

Proof. See Appendix 5.B. □

Defining

$$\boldsymbol{F}_U^{(k)}(x) = \begin{pmatrix} \boldsymbol{\Lambda}(x) - k\boldsymbol{\Delta}(\boldsymbol{r}(x)) & \boldsymbol{R}(x) & \boldsymbol{C}_r^{(2)}(x) & \cdots & \boldsymbol{C}_r^{(k-1)}(x) & \boldsymbol{C}_r^{(k)}(x) \\ \boldsymbol{0} & \boldsymbol{\Lambda}(x) - (k-1)\boldsymbol{\Delta}(\boldsymbol{r}(x)) & \boldsymbol{R}(x) & \cdots & \boldsymbol{C}_r^{(k-2)}(x) & \boldsymbol{C}_r^{(k-1)}(x) \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Lambda}(x) - \boldsymbol{\Delta}(\boldsymbol{r}(x)) & \boldsymbol{R}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Lambda}(x) \end{pmatrix}$$

we get by Van Loan that

$$\prod_t^T (\boldsymbol{I} + \boldsymbol{F}_U^{(k)}(x)\,\mathrm{d}x) = \begin{pmatrix} * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k-1)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k-2)}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(1)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{P}(t,T) \end{pmatrix}. \tag{5.5.6}$$

From these results, we can derive a number of classical results. Differentiation of (5.5.6) gives

$$\begin{pmatrix} * & * & * & * & \cdots & * & \frac{\partial}{\partial t}\boldsymbol{V}_r^{(k)}(t) \\ * & * & * & * & \cdots & * & \frac{\partial}{\partial t}\boldsymbol{V}_r^{(k-1)}(t) \\ * & * & * & * & \cdots & * & \frac{\partial}{\partial t}\boldsymbol{V}_r^{(k-2)}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ * & * & * & * & \cdots & * & \frac{\partial}{\partial t}\boldsymbol{V}_r^{(1)}(t) \\ * & * & * & * & \cdots & * & \frac{\partial}{\partial t}\boldsymbol{P}(t,T) \end{pmatrix} = -\boldsymbol{F}_U^{(k)}(t) \prod_t^T (\boldsymbol{I} + \boldsymbol{F}_U^{(k)}(x)\ \mathrm{d}x)$$

$$= \quad - \begin{pmatrix} \boldsymbol{\Lambda}(t) - k\boldsymbol{\Delta}(\boldsymbol{r}(t)) & \boldsymbol{R}(t) & \boldsymbol{C}_r^{(2)}(t) & \cdots & \boldsymbol{C}_r^{(k-1)}(t) & \boldsymbol{C}_r^{(k)}(t) \\ \boldsymbol{0} & \boldsymbol{\Lambda}(t) - (k-1)\boldsymbol{\Delta}(\boldsymbol{r}(t)) & \boldsymbol{R}(t) & \cdots & \boldsymbol{C}_r^{(k-2)}(t) & \boldsymbol{C}_r^{(k-1)}(t) \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Lambda}(t) - \boldsymbol{\Delta}(\boldsymbol{r}(t)) & \boldsymbol{R}(t) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{\Lambda}(t) \end{pmatrix}$$

$$\times \begin{pmatrix} * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k-1)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(k-2)}(t) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & * & * & \cdots & * & \boldsymbol{V}_r^{(1)}(t) \\ * & * & * & * & \cdots & * & \boldsymbol{P}(t,T) \end{pmatrix}.$$

We then obtain the following differential equation by only considering the first row times the last column.

Theorem 5.5.8. *The matrix of reduced partial state-wise higher order moments satisfies the system of differential equations, for $k \in \mathbb{N}_0$,*

$$\frac{\partial}{\partial s}\boldsymbol{V}_r^{(k)}(t) = (k\boldsymbol{\Delta}(\boldsymbol{r}(t)) - \boldsymbol{\Lambda}(t))\boldsymbol{V}_r^{(k)}(t) - \boldsymbol{R}(t)\boldsymbol{V}_r^{(k-1)}(t) - \sum_{i=2}^{k}\boldsymbol{C}_r^{(i)}(t)\boldsymbol{V}_r^{(k-i)}(t),$$

with terminal condition $\boldsymbol{V}_r^{(k)}(T) = \mathbb{1}_{(k=0)}\boldsymbol{I}$.

Remark 5.5.9. A martingal-based proof for the corresponding (unreduced) state-wise moments, $k!\boldsymbol{V}_r^{(k)}(t)\boldsymbol{e}$, can be found in Norberg (1995b). △

Example 5.5.10 (Independence continued)**.** We can continue our decompositions from the independence case of Example 5.5.2 to reserves and higher-order moments. Indeed, since

$$\begin{aligned} \boldsymbol{\Lambda}_b(u) \oplus \boldsymbol{\Lambda}_r(u) - k\boldsymbol{\Delta}(\boldsymbol{e}\otimes\boldsymbol{r}(u)) &= \boldsymbol{\Lambda}_b(u)\otimes\boldsymbol{I} + \boldsymbol{I}\otimes(\boldsymbol{\Lambda}_r(u) - k\boldsymbol{\Delta}(\boldsymbol{r}(u))) \\ &= \boldsymbol{\Lambda}_b(u) \oplus (\boldsymbol{\Lambda}_r(u) - k\boldsymbol{\Delta}(\boldsymbol{r}(u)), \end{aligned}$$

we get from (5.2.12) that

$$\begin{aligned} &\prod_s^t (\boldsymbol{I} + (\boldsymbol{\Lambda}_b(u) \oplus \boldsymbol{\Lambda}_r(u) - k\boldsymbol{\Delta}(\boldsymbol{e}\otimes\boldsymbol{r}(u))\,\mathrm{d}u)) \\ &\quad= \prod_s^t (\boldsymbol{I} + \boldsymbol{\Lambda}_b(u)\,\mathrm{d}u) \otimes \prod_s^t (\boldsymbol{I} + (\boldsymbol{\Lambda}_r(u) - k\boldsymbol{\Delta}(\boldsymbol{r}(u))\,\mathrm{d}u) \\ &\quad= \prod_s^t (\boldsymbol{I} + \boldsymbol{\Lambda}_b(u)\,\mathrm{d}u) \otimes \boldsymbol{D}^{(k)}(s,t) \end{aligned}$$

Thus, each diagonal block element can be computed using these representations when setting up the matrix $\boldsymbol{F}_U$ for the computation of these higher order moments.

In particular, for partial state-wise reserves (i.e. $k = 1$), we obtain a more direct expression. Assuming that the initial biometric state is $i \in E_b$, the terminal $j \in E_b$ and that the initial distribution of the fitted interest rate phase-type distribution is

$\boldsymbol{\pi}$. Then

$$\begin{aligned}
V_{ij}(t,T) &= (\boldsymbol{e}_i'\otimes\boldsymbol{\pi})\int_t^T\left(\prod_t^x(\boldsymbol{I}+\boldsymbol{\Lambda}_b(u)\,\mathrm{d}u)\otimes\boldsymbol{D}(t,x)\right)(\boldsymbol{R}(x)\otimes\boldsymbol{I})\\
&\quad\times\left(\prod_x^T(\boldsymbol{I}+\boldsymbol{\Lambda}_b(u)\,\mathrm{d}u)\otimes\prod_x^T(\boldsymbol{I}+\boldsymbol{\Lambda}_r(u)\,\mathrm{d}u)\right)\mathrm{d}x\,(\boldsymbol{e}_j\otimes\boldsymbol{e})\\
&= \int_t^T\boldsymbol{\pi}\boldsymbol{D}(t,x)\boldsymbol{e}\boldsymbol{e}_i'\boldsymbol{P}_b(t,x)\boldsymbol{R}(x)\boldsymbol{P}_b(x,T)\boldsymbol{e}_j\,\mathrm{d}x\\
&= \int_t^T\mathrm{E}^{\mathbb{Q}}\left(\mathrm{e}^{-\int_t^T r_{X_r(u)}(u)\,\mathrm{d}u}\,\middle|\,\mathcal{F}(t)\right)\boldsymbol{e}_i'\boldsymbol{P}_b(t,x)\boldsymbol{R}(x)\boldsymbol{P}_b(x,T)\boldsymbol{e}_j\,\mathrm{d}x,
\end{aligned}$$

which is consistent with similar expressions obtained in Norberg (1995a). ◦

5.5.4 Equivalence premium

Assume that $\boldsymbol{R}(t)=\boldsymbol{R}(t;\theta)$ such that θ is a parameter of either $\boldsymbol{B}(t)$ and/or $\boldsymbol{\Delta}(\boldsymbol{b}(t))$ only. Hence, θ could, e.g., be a premium rate in state 1 or a transition payment between some states. We then write $\boldsymbol{V}(t)=\boldsymbol{V}(t;\theta)$ so that

$$\boldsymbol{V}(t;\theta)=\int_t^T\prod_t^u(\boldsymbol{I}+[\boldsymbol{\Lambda}(s)-\boldsymbol{\Delta}(\boldsymbol{r}(s))]\mathrm{d}s)\boldsymbol{R}(u;\theta)\prod_u^T(\boldsymbol{I}+\boldsymbol{\Lambda}(s)\mathrm{d}s)\mathrm{d}u.$$

If the interest rates satisfy $\boldsymbol{\Delta}(\boldsymbol{r}(s))\geq\boldsymbol{0}$, then $\boldsymbol{\Lambda}(s)-\boldsymbol{\Delta}(\boldsymbol{r}(s))$ is a sub-intensity matrix, so that $\prod_t^u(\boldsymbol{I}+[\boldsymbol{\Lambda}(s)-\boldsymbol{\Delta}(\boldsymbol{r}(s))]\mathrm{d}s)$ is a sub-probability matrix, i.e.

$$0\leq\prod_t^u(\boldsymbol{I}+[\boldsymbol{\Lambda}(s)-\boldsymbol{\Delta}(\boldsymbol{r}(s))]\mathrm{d}s)\boldsymbol{e}\leq\boldsymbol{e}.$$

If $\boldsymbol{R}(\cdot;\theta)$ is continuously differentiable and $\boldsymbol{\Lambda}$ and $\boldsymbol{r}$ are continuous, then by Leibniz' integral rule

$$\frac{\partial}{\partial\theta}\boldsymbol{V}(t;\theta)=\int_t^T\prod_t^u(\boldsymbol{I}+[\boldsymbol{\Lambda}(s)-\boldsymbol{\Delta}(\boldsymbol{r}(s))]\mathrm{d}s)\frac{\partial}{\partial\theta}\boldsymbol{R}(u;\theta)\prod_u^T(\boldsymbol{I}+\boldsymbol{\Lambda}(s)\mathrm{d}s)\mathrm{d}u.$$

Hence we get from the Van Loan formula (5.2.9),

$$\prod_t^T\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{\Lambda}(u)-\boldsymbol{\Delta}(\boldsymbol{r}(u)) & \frac{\partial}{\partial\theta}\boldsymbol{R}(u;\theta)\\ \boldsymbol{0} & \boldsymbol{\Lambda}(u)\end{pmatrix}\mathrm{d}u\right)=\begin{pmatrix}\boldsymbol{D}(t,T) & \frac{\partial}{\partial\theta}\boldsymbol{V}(t;\theta)\\ \boldsymbol{0} & \boldsymbol{P}(t,T)\end{pmatrix}. \tag{5.5.7}$$

Remark 5.5.11. Similar kinds of derivatives as those of (5.5.7) are considered in Kalashnikov and Norberg (2003), where differential equations for reserves concerning valuation elements and payments are derived. The formulas presented here may thus be seen as corresponding matrix representations. △

If state $i \in E$ is the starting state, we can formulate the equivalence principle by finding the θ that solves

$$V_i^{Th}(0;\theta) = \boldsymbol{e}_i'\boldsymbol{V}(0;\theta)\boldsymbol{e} = 0$$

using Newton's method,

$$\theta_{n+1} = \theta_n - \frac{\boldsymbol{e}_i'\boldsymbol{V}(0;\theta)\boldsymbol{e}}{\boldsymbol{e}_i'\boldsymbol{V}_\theta(0;\theta)\boldsymbol{e}},$$

where $\boldsymbol{V}_\theta$ denotes the partial derivative wrt. θ. For example, if θ is a constant premium (rate) such that

$$\boldsymbol{R}_\theta(t;\theta) = \boldsymbol{A}(t),$$

i.e. a matrix function not depending on θ, then $\boldsymbol{V}_\theta(t;\theta) = \boldsymbol{V}_\theta(t)$ will not depend on θ either, so we conclude that the map $\theta \mapsto V_i^{Th}(t;\theta)$ is linear (for fixed t), so that in particular

$$V_i^{Th}(0;\theta) = a\theta + b$$

for some constants a, b. Then b can be computed from $b = V_i^{Th}(0;0) = \boldsymbol{e}_i'\boldsymbol{V}(0;0)\boldsymbol{e}$ and $a = \boldsymbol{e}_i'\boldsymbol{V}_\theta(0;0)\boldsymbol{e}$. Hence, Newton's method converges in one iteration, and the θ which fulfills the equivalence principle is given by

$$\theta = -\frac{\boldsymbol{e}_i'\boldsymbol{V}(0;0)\boldsymbol{e}}{\boldsymbol{e}_i'\boldsymbol{V}_\theta(0;0)\boldsymbol{e}}. \tag{5.5.8}$$

Hence, this formula can compute the equivalence premium if it is assumed to be (piecewise) constant over time, which is often the case in practical examples. However, the formulation in terms of derivatives is usually not seen, with Kalashnikov and Norberg (2003, (3.5)) being one of few exceptions. If the constancy assumption is not satisfied, a parametrized expression in terms of θ can be calculated by Newton's method.

5.5.5 Distributions of future payments based on reduced moments

In this section, we briefly comment on the implementation of the Gram-Charlier series for the density and distribution functions based on reduced moments, following along the lines of Bladt, Asmussen, and Steffensen (2020); for an approach based on PDEs and integral equations (though not implemented numerically), we refer to Norberg (2005, Section 5).

The goal is to approximate the distribution of

$$X = \int_0^T \mathrm{e}^{-\int_0^x r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}B(x)$$

using a Gram-Charlier series expansion. In Bladt, Asmussen, and Steffensen (2020), it was shown that under suitable regularity conditions, the density f for X can be approximated by

$$f(x) \approx f^*(x) \sum_{n=0}^{N} c_n p_n(x),$$

where f^* is a reference density, $p_n(x)$ an orthonormal basis of polynomials for Hilbert space $L^2(f^*)$, and $c_n = \mathrm{E}(p_n(X))$. The reference distribution f^* can be chosen arbitrarily as long as $f/f^* \in L^2(f^*)$. Hence it is advisable to choose f^* as close to f as possible.

For a given reference density f^*, the polynomials

$$q_n(x) = \begin{vmatrix} a_0 & \cdots & a_{n-1} & 1 \\ a_1 & \cdots & a_n & x \\ & & \ddots & \\ a_n & \cdots & a_{2n-1} & x^n \end{vmatrix},$$

where

$$a_n = \int_a^b x^n f^*(x)\,\mathrm{d}x, \quad n = 0, 1, ...$$

defines an orthogonal basis for Hilbert space $L^2(f^*)$ with inner product

$$\langle g, h \rangle = \int_a^b g(x)h(x)f^*(x)\,\mathrm{d}x.$$

With the Hankel determinants

$$A_{-1} = 1, \qquad A_n = \begin{vmatrix} a_0 & \cdots & a_{n-1} & a_n \\ a_1 & \cdots & a_n & a_{n+1} \\ & & \ddots & \\ a_n & \cdots & a_{2n-1} & a_{2n} \end{vmatrix}, n = 0, 1, ...$$

it can then be shown that

$$p_n(x) = \frac{q_n(x)}{\sqrt{A_{n-1}A_n}}, \quad n = 0, 1, ...$$

is an orthonormal basis (ONB) in $L^2(f^*)$. Also, it is immediate that

$$c_n = \mathrm{E}(p_n(X)) = \frac{1}{\sqrt{A_{n-1}A_n}} \begin{vmatrix} a_0 & \cdots & a_{n-1} & 1 \\ a_1 & \cdots & a_n & \mathrm{E}(x) \\ & & \ddots & \\ a_n & \cdots & a_{2n-1} & \mathrm{E}(x^n) \end{vmatrix}.$$

If f^* is chosen to be the standard normal distribution, the corresponding polynomials p_n are the (probabilists) Hermite polynomials. While the Hermite polynomials were used in Bladt, Asmussen, and Steffensen (2020) up to very high orders, their use in the following example fails already at low orders. This is likely caused by the tail of the normal distribution being too light. We propose a class of reference distributions based on a shifted beta distribution closely related to the Jacobi Polynomials as an alternative. This distribution will have finite support but a much heavier tail. Finite support is usually not a problem in a life insurance context.

Define a reference distribution f^* with support on a finite interval $[a,b]$ by

$$f^*(x) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}(b-a)^{-\alpha-\beta-1}(b-y)^{\alpha}(y-a)^{\beta}, \quad x \in [a,b], \ \alpha,\beta > -1.$$

Thus we need to find an orthonormal basis for $L^2(f^*)$. The starting point is the weight function

$$w^{\alpha,\beta}(x) = (1-x)^{\alpha}(1+x)^{\beta}.$$

The space $L^2(w)$ has an orthogonal basis of Jacobi polynomials given by

$$q_n^{(\alpha,\beta)}(x) = \frac{(\alpha+1)_n}{n!}\sum_{k=0}^{n}\frac{(\alpha+\beta+1+n)_k(-n)_k}{(\alpha+1)_k k!}\left(\frac{1-x}{2}\right)^k,$$

where $(a)_n = a(a+1)\cdots(a+n-1)$ denotes the Pochammer symbol.

By normalizing the weight function into a density on $[-1,1]$ and then transforming it into a density on $[a,b]$, we obtain an ONB for f^* of polynomials given by

$$p_n^{\alpha,\beta}(x) = \sqrt{\frac{n!(2n+\alpha+\beta+1)(\alpha+\beta+1)_n}{(\alpha+1)_n(\beta+1)_n(\alpha+\beta+1)}}\, q_n^{(\alpha,\beta)}\left(\frac{2x-a-b}{b-a}\right).$$

So for given a, b, we need to compute

$$\begin{aligned} c_n &= \mathrm{E}\left(p_n^{(\alpha,\beta)}(X)\right) \\ &= \sqrt{\frac{n!(2n+\alpha+\beta+1)(\alpha+\beta+1)_n}{(\alpha+1)_n(\beta+1)_n(\alpha+\beta+1)}}\, \mathrm{E}\left(q_n^{(\alpha,\beta)}\left(\frac{2X-a-b}{b-a}\right)\right). \end{aligned}$$

Here

$$\begin{aligned} &\mathrm{E}\left(q_n^{(\alpha,\beta)}\left(\frac{2X-a-b}{b-a}\right)\right) \\ &= \frac{(\alpha+1)_n}{n!}\sum_{k=0}^{n}\frac{(\alpha+\beta+1+n)_k(-n)_k}{(\alpha+1)_k}\frac{1}{k!}\,\mathrm{E}\left(\left(\frac{1-(2X-a-b)/(b-a)}{2}\right)^k\right), \end{aligned}$$

where the inner expectation is computed as

$$\frac{1}{k!}\,\mathrm{E}\left(\left(\frac{1-(2X-a-b)/(b-a)}{2}\right)^k\right) = \frac{1}{(b-a)^k}\sum_{i=0}^{k}(-1)^i\frac{b^{k-i}}{(k-i)!}\frac{\mathrm{E}(X^i)}{i!}.$$

Finally, the approximation is then given by

$$f(x) \approx f^*(x) \sum_{n=0}^{N} c_n p_n^{(\alpha,\beta)}(x). \tag{5.5.9}$$

Concerning the corresponding distribution function, we integrate the above equation to obtain

$$\begin{aligned} F(y) \quad &\approx \quad F^*(y) - \frac{b-a}{4}\left(1-\left(\frac{2y-a-b}{b-a}\right)^2\right) f^*(y) \\ &\times \sum_{n=1}^{N} c_n \sqrt{\frac{1}{n}\frac{(2+\alpha+\beta)(\alpha+\beta+3)}{(1+\alpha)(1+\beta)(\alpha+\beta+n+1)(\alpha+\beta+n+2)}} \times \\ &\quad p_{n-1}^{(\alpha+1,\beta+1)}\left(\frac{2y-a-b}{b-a}\right). \end{aligned}$$

Hence, these formulas can be used to approximate the density and distribution via these Jacobi types of polynomials.

5.6 Numerical Example

We now present a numerical example based on Example 5.5.2, where interest rates and biometric risk are assumed independent, where we carry over the estimation of interest transition rates from the calibrated bond prices of Section 5.4.

Consider the numerical example of Buchardt and Møller (2015) as the model for the biometric risk and corresponding life insurance contract. That is, the states of the insured X_b are modeled as a time-inhomogeneous Markov jump process taking values $E_b = \{1, 2, 3\}$, the three-state disability model depicted in Figure 5.7.

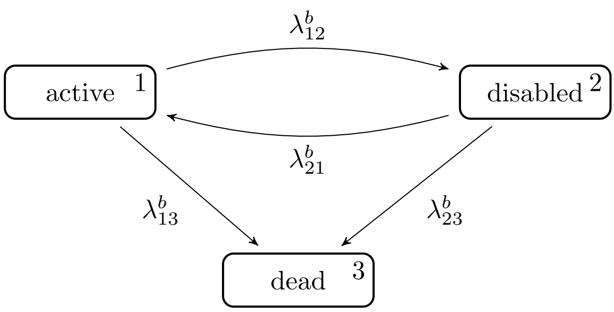

Figure 5.7: *The classic three-state disability model with reactivation*

We consider a 40-year-old male today (at time 0) with a retirement age of 65 and the following life insurance contract:

- A disability annuity of rate 1 while disabled until the retirement of age 65.

- A life annuity of rate 1 while alive until the retirement of age 65.
- A constant premium rate θ paid while active until the retirement of age 65, priced under the equivalence principle at time 0.

The maximum contract time is $T = 70$, corresponding to a maximum age of 110 years. The transition rates are given by

$$\begin{aligned}
\lambda_{12}^b(s) &= \left(0.0004 + 10^{4.54+0.06(s+40)-10}\right)1_{(s\leq 25)},\\
\lambda_{21}^b(s) &= \left(2.0058e^{-0.117(s+40)}\right)1_{(s\leq 25)},\\
\lambda_{13}^b(s) &= 0.0005 + 10^{5.88+0.038(s+40)-10},\\
\lambda_{23}^b(s) &= \lambda_{13}^b(s)\big(1 + 1_{(s\leq 25)}\big).
\end{aligned}$$

The payment matrices for this product combination corresponds to having $\boldsymbol{B}(t) = \boldsymbol{\Lambda}^1(t) = \boldsymbol{0}$, and

$$\boldsymbol{b}(t;\theta) = \begin{cases} (\theta, 1, 0), & \text{for } t \leq 25 \\ (1, 1, 0), & \text{for } t > 25 \end{cases}.$$

For the stochastic interest rate model, we take the fitted bond prices from Example 5.4.1 with $p = 4$ phases, so that the interest rates are given as $r(t) = r_{X_r(t)}$, with

$$\boldsymbol{r} = (0.025, 0.050, 0.075, 0.100),$$

and where X_r is a time-homogeneous Markov jump process taking values the finite state space $E_r = \{1, 2, 3, 4\}$ with initial distribution $\boldsymbol{\pi} = (1, 0, 0, 0)$ and transition intensity matrix

$$\boldsymbol{\Lambda}_r = \begin{pmatrix} -0.25 & 0.22 & 0.01 & 0 \\ 0.14 & -1.11 & 0.75 & 0.18 \\ 0.06 & 0.29 & -0.63 & 0.2 \\ 0.09 & 0.22 & 0.65 & -1.05 \end{pmatrix} + \boldsymbol{\Delta}(\boldsymbol{r}).$$

We then determine the equivalence premium θ using the method outlined in Section 5.5.4. This is explicit on the form (5.5.8) due to $\boldsymbol{b}(t;\cdot)$ being affine (for fixed t), and we get $\theta = 0.1583467$. This is almost three times lower than the premium rate obtained when pricing with a constant first-order interest rate of 1% as in Buchardt and Møller (2015), which makes sense since the present interest rate model always gives interest rates above this level.

We then calculate moments of up to order 20 of the present value of future payments to approximate its density and distribution function via Gram-Charlier expansions based on the (shifted) Jacobi polynomials, as outlined in Section 5.5.5. The parameters used in the procedure are shown in Table 5.1, and the resulting

Parameter	α	β	a	b
Value	1	0.05	-3	70

Table 5.1: *Parameters for the Gram-Charlier implementation with (shifted) Jacobi polynomials.*

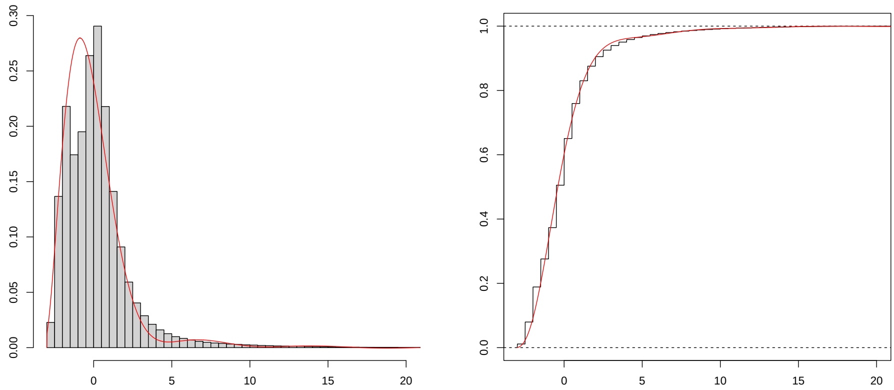

Figure 5.8: *Left: Density approximation based on 20 moments and a histogram based on* 1,000,000 *simulations. Right: Distribution function approximation based on the same 20 moments and the empirical distribution function from the same simulations.*

density and distribution function are shown in Figure 5.8. From the fitted distribution function, one may compute different quantities of interest, e.g., quantiles of the present value. In Table 5.2, we show various quantiles based on the empirical (simulated) distribution function and the approximated distribution function based on 20 moments.

	Quantile			
	95%	97%	99%	99.5%
Empirical	3.51	5.51	9.51	12.01
Moment-based	3.13	5.54	8.89	12.63

Table 5.2: *Selected quantiles of the present value based on the empirical distribution of* 1,000,000 *simulations and based on the Gram-Charlier approximation based on 20 moments.*

5.A Modified EM algorithm for phase-type fitting with fixed exit rate vector

First, we consider the case we want to fit a phase-type distribution with parameters $(\boldsymbol{\pi}, \boldsymbol{T}(x))$ to data $y_1, ..., y_N$. Here the data are positive real numbers which are thought of as the time until absorption of the underlying Markov process with intensity matrix

$$\begin{pmatrix} \boldsymbol{T}(x) & \boldsymbol{t}(x) \\ \boldsymbol{0} & 0 \end{pmatrix}.$$

We will assume throughout that $\boldsymbol{T}(x) = \boldsymbol{T}$, i.e. the Markov process is time-homogeneous. This presents no restriction as long as the interest rate process to be approximated is assumed to be stationary.

If, additionally to the absorption times, we could observe the full trajectories of the Markov process until absorption, then the estimation would be an easy task. In this case, for $i \neq j$

$$\hat{\pi}_i = \frac{B_i}{N}, \hat{t}_{ij} = \frac{N_{ij}}{Z_i}, \hat{t}_i = \frac{N_i}{Z_i} \tag{5.A.1}$$

whereas $\hat{t}_{ii} = -\hat{t}_i - \sum_{j \neq i} \hat{t}_{ij}$. Here B_i denotes the number of processes starting in state i, N_{ij} the number of transitions from i to j in all processes, N_i the number of processes that exits to the absorbing state from state i and Z_i the total time all processes spend in state i.

In the case of incomplete data, where only absorption times are observed, the EM–algorithm can be employed. The idea is to replace the unobserved sufficient statistics B_i, N_{ij}, N_i and Z_i by the their conditional expectations given data, i.e. $\mathrm{E}(B_i | Y = y)$ etc. The EM-algorithm then alternates between computing these conditional expected values (E-step) and plugging them into (5.A.1) as a substitute, thereby generating new parameters.

To perform maximization under the constraint $\boldsymbol{t}(x) = \boldsymbol{t} = \boldsymbol{r}$, we see that this can be achieved simply by removing the update $\hat{t}_i = \frac{N_i}{Z_i}$ from the EM-algorithm, so that the M-step amounts to

$$\hat{\pi}_i = \frac{B_i}{N}, \hat{t}_{ij} = \frac{N_{ij}}{Z_i},\ i \neq j,\ \hat{t}_{ii} = -\hat{r}_i - \sum_{j \neq i} \hat{t}_{ij}. \tag{5.A.2}$$

Fitting a PH distribution to a theoretical distribution is done by approximating the theoretical distribution into a histogram. Hence data will be the discretization points, and the density values will be the corresponding weights. For further details on the EM-algorithm, we refer to Asmussen, Nerman, and Olsson (1996) or Albrecher, Bladt, and Yslas (2022).

5.B Proofs

Proof of Theorem 5.5.3. First, we notice that, for $i, j \in E$,

$$\begin{aligned} &V_{ij}(s,t) \\ &= \sum_{k \in E} \mathrm{E}\left(\mathbb{1}_{(X(t)=j)} \int_s^t \mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u} \,\mathrm{d}B(x) \middle| X(s) = i \right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k\in E} \int_s^t \mathrm{E}\left(\mathbb{1}_{(X(t)=j)} \mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}B(x) \Big|\, X(s)=i \right) \\
&= \sum_{k\in E} \int_s^t \mathrm{E}\left(\mathrm{E}\left(\mathbb{1}_{(X(t)=j)} \mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}B(x) \Big|\, \mathcal{F}_x \right) \Big|\, X(s)=i \right) \\
&= \sum_{k\in E} \int_s^t \mathrm{E}\left(\mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{E}\left(\mathbb{1}_{(X(t)=j)}\, \mathrm{d}B(x) \big|\, \mathcal{F}_x \right) \Big|\, X(s)=i \right).
\end{aligned}
$$

But on the event $(X(x)=k)$,

$$
\mathrm{E}\left(\mathbb{1}_{(X(t)=j)}\, \mathrm{d}B(x) \big|\, \mathcal{F}_x \right) = b_k(x)\mathrm{d}x\ p_{kj}(x,t) + \sum_{\substack{\ell\in E\\ \ell\neq k}} b_{k\ell}(x)\lambda^1_{k\ell}(x)\mathrm{d}x\ p_{\ell j}(x,t)
$$

so

$$
\begin{aligned}
&V_{ij}(s,t) \\
&= \sum_{k\in E} \int_s^t p_{kj}(x,t)\, \mathrm{E}\left(\mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u} \Big|\, X(s)=i \right) b_k(x)\,\mathrm{d}x \\
&+ \sum_{k\in E} \int_s^t \mathrm{E}\left(\mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u} \left(\sum_{\substack{\ell\in E\\ \ell\neq k}} b_{k\ell}(x)\lambda^1_{k\ell}(x)\,\mathrm{d}x\ p_{\ell j}(x,t) \right) \Bigg|\, X(s)=i \right) \\
&= \sum_{k\in E} \int_s^t p_{kj}(x,t)\, \mathrm{E}\left(\mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u} \Big|\, X(s)=i \right) b_k(x)\,\mathrm{d}x \\
&+ \sum_{k\in E} \int_s^t \left(\sum_{\substack{\ell\in E\\ \ell\neq k}} b_{k\ell}(x)\lambda^1_{k\ell}(x)\ p_{\ell j}(x,t) \right) \mathrm{E}\left(\mathbb{1}_{(X(x)=k)} \mathrm{e}^{-\int_s^x r_{X(u)}(u)\,\mathrm{d}u} \Big|\, X(s)=i \right) \mathrm{d}x \\
&= \int_s^t \sum_{k\in E} D_{ik}(s,x) b_k(x) p_{kj}(x,t)\,\mathrm{d}x + \int_s^t \sum_{\substack{k,\ell\in E\\ \ell\neq k}} D_{ik}(s,x) b_{k\ell}(x)\lambda^1_{k\ell}(x) p_{\ell j}(x,t)\,\mathrm{d}x.
\end{aligned}
$$

In matrix form this amounts to (5.5.5). □

Proof of Theorem 5.5.5. Using that the product integral satisfies Kolmogorov's forward and backward equations, we get that

$$
\begin{aligned}
\begin{pmatrix} \frac{\partial}{\partial s}\boldsymbol{D}(s,t) & \frac{\partial}{\partial s}\boldsymbol{V}(s,t) \\ \boldsymbol{0} & \frac{\partial}{\partial s}\boldsymbol{P}(s,t) \end{pmatrix}
&= \frac{\partial}{\partial s} \prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{\Lambda}(u) - \boldsymbol{\Delta}(\boldsymbol{r}(u)) & \boldsymbol{R}(u) \\ \boldsymbol{0} & \boldsymbol{\Lambda}(u) \end{pmatrix} \mathrm{d}u \right) \\
&= - \begin{pmatrix} \boldsymbol{\Lambda}(s) - \boldsymbol{\Delta}(\boldsymbol{r}(s)) & \boldsymbol{R}(s) \\ \boldsymbol{0} & \boldsymbol{\Lambda}(s) \end{pmatrix} \begin{pmatrix} \boldsymbol{D}(s,t) & \boldsymbol{V}(s,t) \\ \boldsymbol{0} & \boldsymbol{P}(s,t) \end{pmatrix},
\end{aligned}
$$

from which Thiele's differential equation can be pulled out from the upper right corner of each side of the equation. □

Proof of Theorem 5.5.7. Write

$$
\begin{aligned}
&\left(\int_t^T \mathrm{e}^{-\int_t^x r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(x)\right)^k \\
&\quad= \int_t^T \cdots \int_t^T \mathrm{e}^{-\int_t^{x_1} r_{X(u)}(u)\,\mathrm{d}u} \cdots \mathrm{e}^{-\int_t^{x_k} r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(x_k)\cdots\,\mathrm{d}A(x_1).
\end{aligned}
\tag{5.B.1}
$$

Now

$$
\mathrm{d}A(t) = b_{X(t)}\,\mathrm{d}t + b_{X(t-)j}(t)\,\mathrm{d}N_{X(t-)j}(t),
$$

and assume that $s \in [t, T]$ is a point of increase for the counting process $x \to N_{ab}(x)$ which trigger lump sum payments. Then in the computation of the above integral, there will be jump contributions at time s, where any number $m \in \{1, 2, ..., k\}$ of the variables $x_1, ..., x_k$ may be equal to s, say $x_{i_1} = \cdots x_{i_m} = s$. We can pick m out of the k variables in $\binom{k}{m} = k!/(m!(k-m)!)$ ways. If m variables coincide at the jump time s, then a contribution of $b_{ab}(s)^m$ is added. Hence only looking at jump coincidences, i.e. $m \geq 2$, the contribution to the integral (5.B.1) is

$$
\begin{aligned}
&\sum_{m=2}^{k} \binom{k}{m} \int_t^T \mathrm{e}^{-m\int_t^s r_{X(u)}(u)\,\mathrm{d}u} b_{ab}(s)^m \times \\
&\left(\int_s^T \cdots \int_s^T \mathrm{e}^{-\int_t^{x_{m+1}} r_{X(u)}(u)\,\mathrm{d}u} \cdots \mathrm{e}^{-\int_t^{x_k} r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(x_k)\cdots\,\mathrm{d}A(x_{m+1})\right)\mathrm{d}N_{ab}(s).
\end{aligned}
$$

Indeed, since there are precisely m coincidences, the remaining integrals must start from $s+ = s$; otherwise, the integration intervals would contain s as well. Changing the lower limits of the integrals appearing in the exponentials, we can further rewrite the expression as

$$
\begin{aligned}
&\sum_{m=2}^{k} \binom{k}{m} \int_t^T \mathrm{e}^{-k\int_t^s r_{X(u)}(u)\,\mathrm{d}u} b_{ab}(s)^m \times \\
&\left(\int_s^T \cdots \int_s^T \mathrm{e}^{-\int_s^{x_{m+1}} r_{X(u)}(u)\,\mathrm{d}u} \cdots \mathrm{e}^{-\int_s^{x_k} r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(x_k)\cdots\,\mathrm{d}A(x_{m+1})\right)\mathrm{d}N_{ab}(s).
\end{aligned}
\tag{5.B.2}
$$

Taking conditional expectation $\mathrm{E}(\mathbb{1}_{(X(T)=j)} \cdot | X(t) = i)$ of (5.B.2), we get

$$
\begin{aligned}
&\sum_{m=2}^{k} \binom{k}{m} \int_t^T \mathrm{E}\Bigg(\mathbb{1}_{(X(T)=j)} \mathrm{e}^{-k\int_t^s r_{X(u)}(u)\,\mathrm{d}u} b_{ab}(s)^m \times \\
&\left(\int_s^T \cdots \int_s^T \mathrm{e}^{-\int_s^{x_{m+1}} r_{X(u)}(u)\,\mathrm{d}u} \cdots \mathrm{e}^{-\int_s^{x_k} r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(x_k)\cdots\,\mathrm{d}A(x_{m+1})\right)\mathrm{d}N_{ab}(s) \\
&\hspace{28em}\Bigg| X(t) = i\Bigg) \\
&= \sum_{m=2}^{k} \binom{k}{m} \int_t^T \mathrm{E}\left(\mathbb{1}_{(X(T)=j)} Y\, Z\,\mathrm{d}N_{ab}(s)\middle|\, X(t) = i\right),
\end{aligned}
\tag{5.B.3}
$$

where $Y = \mathrm{e}^{-k\int_t^s r_{X(u)}(u)\,\mathrm{d}u} b_{ab}(s)^m$ and

$$Z = \int_s^T \cdots \int_s^T \mathrm{e}^{-\int_s^{x_{m+1}} r_{X(u)}(u)\,\mathrm{d}u} \cdots \mathrm{e}^{-\int_s^{x_k} r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(x_k) \cdots \mathrm{d}A(x_{m+1}).$$

Further conditioning on a lump sum triggering event at time s, caused by $(X(s) = b,\ X(s-) = a)$, the probability of which is $d_{ab}(s)\,\mathrm{d}s$, and using that $\mathbb{1}_{(X(T)=j)}Z$ and $\mathbb{1}_{(X(s-)=a)}Y$ are conditionally independent given $X(s) = b$, (5.B.3) reduces to

$$\begin{aligned}
&\sum_{m=2}^{k} \binom{k}{m} \int_t^T \mathrm{E}\left(\mathbb{1}_{(X(T)=j)} Z \middle| X(s) = b \right) \mathrm{E}\left(\mathbb{1}_{(X(s)=a)} Y \middle| X(t) = i \right) d_{ab}(s)\,\mathrm{d}s \\
&= \sum_{m=2}^{k} \binom{k}{m} \int_t^T D_{ia}^{(k)}(t,s) d_{ab}(s) b_{ab}(s)^m V_{bj}^{(k-m)}(s,T)\,\mathrm{d}s. \qquad (5.B.4)
\end{aligned}$$

Summing over a and b, and putting (5.B.4) on matrix form (in i, j) this amounts to

$$\sum_{m=2}^{k} \binom{k}{m} \int_t^T \boldsymbol{D}^{(k)}(t,s) \boldsymbol{C}^{(m)}(s) \boldsymbol{V}^{(k-m)}(s,T)\,\mathrm{d}s. \qquad (5.B.5)$$

Now we consider the integral when there are no coincidences. To this end, we rewrite

$$\begin{aligned}
&\left(\int_t^T \mathrm{e}^{-\int_t^x r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(x) \right)^k \\
&= k \int_t^T \mathrm{e}^{-\int_t^x r_{X(u)}(u)\,\mathrm{d}u} \left(\int_x^T \mathrm{e}^{-\int_t^y r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(y) \right)^{k-1} \mathrm{d}A(x) \\
&= k \int_t^T \mathrm{e}^{-k\int_t^x r_{X(u)}(u)\,\mathrm{d}u} \left(\int_x^T \mathrm{e}^{-\int_x^y r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(y) \right)^{k-1} \mathrm{d}A(x)
\end{aligned}$$

Then

$$\begin{aligned}
&V_{ij}^{(k)}(t,T) \\
&= \mathrm{E}\left(\mathbb{1}_{(X(T)=j)} \left(\int_t^T \mathrm{e}^{-\int_t^x r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(x) \right)^k \middle| X(t) = i \right) \qquad (5.B.6) \\
&= k \int_t^T \mathrm{E}\Bigg(\mathbb{1}_{(X(T)=j)} \mathrm{e}^{-k\int_t^x r_{X(u)}(u)\,\mathrm{d}u} \times \qquad (5.B.7) \\
&\qquad \left(\int_x^T \mathrm{e}^{-\int_x^y r_{X(u)}(u)\,\mathrm{d}u}\, \mathrm{d}A(y) \right)^{k-1} \mathrm{d}A(x) \Bigg| X(t) = i \Bigg) \\
&= k \sum_{\ell} \int_t^T \mathrm{E}\Bigg(\mathbb{1}_{(X(x)=\ell)} \mathrm{e}^{-k\int_t^x r_{X(u)}(u)\,\mathrm{d}u} \mathbb{1}_{(X(T)=j)} \times
\end{aligned}$$

$$\mathrm{E}\left(\mathbb{1}_{(X(T)=j)}\left(\int_x^T \mathrm{e}^{-\int_x^y r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(y)\right)^{k-1}\,\mathrm{d}A(x)\,\middle|\,X(x)=\ell\right)\middle|\,X(t)=i\right). \tag{5.B.8}$$

On the event that $(X(x)=\ell)$, the contribution to the expectation of the above integral (5.B.8), where no coincidences are allowed (i.e., the reward at time x from at most one jump and benefit rates), amounts to

$$\begin{aligned}
&\mathrm{E}\left(\mathbb{1}_{(X(T)=j)}\left(\int_x^T \mathrm{e}^{-\int_x^y r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(y)\right)^{k-1}\middle|X(x)=\ell\right)b_\ell(x)\,\mathrm{d}x\\
&+\sum_m \mathrm{E}\left(\mathbb{1}_{(X(T)=j)}\left(\int_x^T \mathrm{e}^{-\int_x^y r_{X(u)}(u)\,\mathrm{d}u}\,\mathrm{d}A(y)\right)^{k-1}\middle|X(x)=m\right)d_{\ell m}(x)b_{\ell m}(x)\mathrm{d}x\\
&=b_\ell(x)V_{\ell j}^{(k-1)}(x,T)\,\mathrm{d}x+\sum_m d_{\ell m}(x)b_{\ell m}(x)V_{mj}^{(k-1)}(x,T)\,\mathrm{d}x,
\end{aligned}$$

and the integral (5.B.8) then equals

$$\begin{aligned}
k\sum_\ell\int_t^T \mathrm{E}\left(1\{X(x)=\ell\}\mathrm{e}^{-k\int_t^x r_{X(u)}(u)\,\mathrm{d}u}\middle|X(t)=i\right)\times\\
\left(b_\ell(x)V_{\ell j}^{(k-1)}(x,T)+\sum_m d_{\ell m}(x)b_{\ell m}(x)V_{mj}^{(k-1)}(x,T)\right)\mathrm{d}x,
\end{aligned}$$

which in matrix form amounts to

$$k\int_t^T \boldsymbol{D}^{(k)}(t,x)\boldsymbol{R}(x)\boldsymbol{V}^{(k-1)}(x,T)\,\mathrm{d}x. \tag{5.B.9}$$

Adding (5.B.5) and (5.B.9) then proves the result. □

Chapter 6

Multivariate higher order moments in multi-state life insurance

This chapter is based on the paper *Ahmad (2022)*.

Abstract

It is well-known that combining life annuities and death benefits introduce opposite effects in payments with respect to the mortality risk on the lifetime of the insured. In a general multi-state framework with multiple product types, such joint effects are less trivial. In this paper, we consider a multivariate payment process in multi-state life insurance, where the components are defined in terms of the same Markovian state process. The multivariate present value of future payments is introduced, and we derive differential equations and product integral representations of its conditional moments and moment generating function. Special attention is given to pair-wise covariances between two present values, where results closely connected to Hattendorff type of results for the variance are derived. The results are illustrated in a numerical example in a disability model.

Keywords: Multi-state life insurance; Multivariate payment process; Dependent risk; Conditional moments; Product integral

6.1 Introduction

In this paper, we extend the results of Bladt, Asmussen, and Steffensen (2020) to multivariate payment processes with components defined in terms of the same multi-state Markov process. Our main contributions are differential equations and product integral representations of higher order moments of the multivariate present value, which are derived via its moment generating function. The main results appear as natural multivariate generalizations to those of Bladt, Asmussen, and Steffensen (2020) and Norberg (1995b), and to some extent also Adékambi and Christiansen (2017), and this is pointed out in a series of remarks throughout the paper. We give special attention to pair-wise covariances between two present values and derive results that reveal close connections to Hattendorff type of results for the variance of a present value.

The paper is motivated by the following. The use of multi-state time–inhomogeneous Markov models in life insurance, dating back to at least Hoem (1969a), have provided a unified mathematical framework to model the random pattern of states of the insured with different kinds of life and health events; see e.g. an overview in Asmussen and Steffensen (2020, Chapter V.2). While these are models for primarily unsystematic biometric risk, one may also use multi-state Markov models to integrate systematic risk as in Norberg (1995a) and Norberg (1999). For valuation and risk management, the present value of future payments is the main quantity of interest, and the conditional expected present value constitute the prospective reserve; this principle relies on diversification of unsystematic risk resulting from the law of large numbers on independent and identically distributed present values. The prospective reserve satisfies the celebrated Thiele's differential equations, cf. Hoem (1969a).

Since there may be a significant risk of future payments deviating from expected values, the insurer adds safety margins to premiums and reserves. While this is usually done implicitly via a first order basis consisting of prudent assumptions on interest and transition rates, an alternative approach is to compute safety margins explicitly via (properties of) the probability distribution of future payments. This is the focal point of Christiansen (2013), where approximations based on the Central Limit Theorem (CLT) are established. Here, the variance of future payments is needed; it is obtained as an integral expression in Hoem (1969a), see also Hoem and Aalen (1978) and Norberg (1991), and Hattendorff's theorem, which is formulated in the multi-state framework by Ramlau-Hansen (1988) and Norberg (1992), provides particularly simple formulae in terms of the associated multivariate counting process and so-called sum at risks. Turning to higher order moments beyond the variance increases the precision in these kinds of approximations; in Norberg (1995b), we have differential equations for all moments of future payments, and in Bladt, Asmussen, and Steffensen (2020), a general matrix-based framework using product integrals is developed to compute said moments, from which densities and distribution functions of future payments are approximated via polynomial expansions. These type of results are extended to the semi-Markovian framework in Adékambi and Christiansen (2017) to allow for duration dependency in payments and transition rates. Other ways of calculating distribution functions can be found in Hesselager and Norberg (1996) and Adékambi and Christiansen (2020), where integral and differential equations for these are derived.

In practice, an insured typically holds a combination of various product types in order to be covered in case of different kinds of life and health events. In the multi-state framework, this is without further notice handled by considering the aggregated payment process, and reserves, higher order moments and probability distributions can be calculated using the methods outlined above. However, since these product types are contingent on events that work in different directions w.r.t. their underlying risks, a decomposition of the total risk into the different risk types is important in pricing and reserving as well as in risk management. In Christiansen (2013), asymptotic safety margins are obtained for premiums and reserves, from which a decomposition of risk types corresponding to the different transitions in the Markov chain is established. In this paper, we take a different point of view; we decompose the payment process into different product types and examine the joint distribution of their future payments by calculating its moments; in particular, we calculate covariance matrices and illustrate how they may be used to approximate joint safety margins via multivariate CLT approximations. The simplest example is the product combination of life annuities and death benefits, which have opposite effects with respect to the mortality risk on the lifetime of the insured. In the general multi-state framework of this paper, more complex and non-trivial interactions can be examined; we give a motivating example in a disability model in the next section.

For systematic risk, the idea of mixing various product types and analysing their interacting effect has already been discussed extensively in the concept of so-called natural hedging in life insurance, where one considers the effects on the reserves with respect to future changes in mortality rates in a portfolio consisting exactly of life annuities and death benefits; see e.g. Cox and Lin (2007). While this concept arose in the survival model at first, an example in the multi-state model have appeared recently in Levantesi and Menzietti (2018) where the authors analyze interactions with life products and long-term care insurance using a disability model. Results from this paper allows us to carry out these kinds of analyzes in a general multi-state framework, however with models for systematic risk restricted to time-inhomogeneous Markov chains on finite state spaces.

The paper is structured as follows. We start out in Section 6.2 with presenting a motivating example of decomposition of payment processes into different types. Section 6.3 then introduces the setup of the paper: the multi-state Markov process governing the state of the insured, the multivariate payment process describing the collection of life insurance contracts held by the insured, and the multivariate present value of their payments. In Section 6.4, we derive product integral representations of the conditional moment generating function of the present value, which is shown to satisfy a system of partial differential equations. In Section 6.5 we use this to derive ordinary differential equations and product integral representations for the conditional moments. Finally, in Section 6.6 we illustrate results of this paper in a numerical example of the motivating example from Section 6.2.

6.2 Motivating example

Before presenting the general setup and main results of the paper, we start out with giving a motivating example of the problems we wish to solve. The setup of the example is a disability model, which serves to illustrate the motivations in the simplest non-trivial multi-state life insurance setting.

Consider a single insured having a deterministic retirement time T with the following product combination:

(1) Death benefit paying S upon death before time T.

(2) Deferred life annuity paying a benefit rate of b while alive, starting from time T.

(3) Disability annuity paying a benefit rate of d while disabled until time T.

The state of the insured is modeled in the disability model with recoveries, as depicted in Figure 6.1.

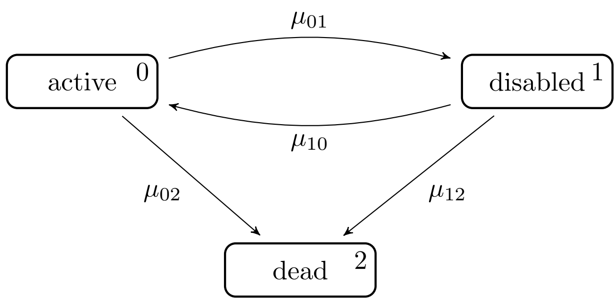

Figure 6.1: *Disability model with recoveries.*

The payment process B giving accumulated benefits from this product design then takes the form

$$\begin{aligned} \mathrm{d}B(t) &= \mathbb{1}_{(Z(t-)=0)}\big(b \cdot \mathbb{1}_{(t\geq T)}\,\mathrm{d}t + S \cdot \mathbb{1}_{(t<T)}\,\mathrm{d}N_{02}(t)\big) \\ &\qquad + \mathbb{1}_{(Z(t-)=1)}\big(d \cdot \mathbb{1}_{(t<T)}\,\mathrm{d}t + b \cdot \mathbb{1}_{(t\geq T)}\,\mathrm{d}t + S \cdot \mathbb{1}_{(t<T)}\,\mathrm{d}N_{12}(t)\big)\,, \\ B(0) &= 0, \end{aligned}$$

where $Z = \{Z(t)\}_{t\geq 0}$ is the process (with state-space according to Figure 6.1) indicating the state of the insured and N_{ij} is the associated counting process counting the number of jumps from state i to j, for $i, j \in \{0, 1, 2\}$, $i \neq j$.

The present value at initiation of the contract (at time 0) of future benefits up to some terminal time point $t > 0$ is then given by

$$U(0,t) = \int_0^t e^{-\int_0^s r(v)\,\mathrm{d}v}\,\mathrm{d}B(s),$$

where $r = \{r(v)\}_{v\geq 0}$ is some deterministic interest rate that accounts for time value of money. Based on a principle of diversification of unsystematic biometric risk, the prospective reserve is given as the expected present value, i.e. $V(0,t) = \mathbb{E}[U(0,t)]$, which satisfies the well-known Thiele's differential equation, cf. Hoem (1969a). This principle relies on the strong law of large numbers on i.i.d. present values (coming from i.i.d. insured with same product combination), $U^1(0,t), U^2(0,t), \ldots$,

$$\frac{1}{N}\sum_{i=1}^{N} U^i(0,t) \xrightarrow{\text{a.s}} \mathbb{E}\big[U^1(0,t)\big], \quad \text{as} \quad N \to \infty.$$

To avoid a significant risk of future payments deviating from expected values, the insurer adds safety margins to the prospective reserve. Assuming an explicit approach based on the probability distribution of the present value, a natural approach is to use approximations based on the Central Limit Theorem (CLT), cf. Christiansen (2013):

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} U^i(0,t) - \mathbb{E}\big[U^1(0,t)\big]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \mathrm{Var}\big[U^1(0,t)\big]\big), \quad \text{as} \quad N \to \infty.$$

We recall that $\mathrm{Var}\big[U^1(0,t)\big]$ can be calculated using multi-state versions of Hattendorff's theorem, cf. Ramlau-Hansen (1988). One may even turn to higher order moments beyond the variance to approximate these distributions further, cf. Norberg (1995b) and Bladt, Asmussen, and Steffensen (2020).

Now, since the different payment types work in different directions w.r.t. the underlying mortality and disability risk, the safety margins on the individual product types are dependent. We are therefore interested in computing joint safety margins that suitably represents decompositions of these type of risks. This we can obtain by decomposing the payment process B into the three types corresponding to (1) – (3) above:

$$\begin{aligned}
B(t) &= B_1(t) + B_2(t) + B_3(t), \\
\mathrm{d}B_1(t) &= S \cdot \mathbb{1}_{(t<T)}\big(\mathbb{1}_{(Z(t-)=0)}\,\mathrm{d}N_{02}(t) + \mathbb{1}_{(Z(t-)=1)}\,\mathrm{d}N_{12}(t)\big)\,, \quad && B_1(0) = 0, \\
\mathrm{d}B_2(t) &= b \cdot \mathbb{1}_{(t\geq T)}\big(\mathbb{1}_{(Z(t-)=0)} + \mathbb{1}_{(Z(t-)=1)}\big)\,\mathrm{d}t, && B_2(0) = 0,
\end{aligned}$$

$$\mathrm{d}B_3(t) = d \cdot \mathbb{1}_{(t<T)} \mathbb{1}_{(Z(t-)=1)}\,\mathrm{d}t, \qquad\qquad B_3(0) = 0,$$

and where the present value of the payment processes B_1, B_2 and B_3 are coupled in a vector: $\boldsymbol{U}(0,t) = (U_1(0,t), U_2(0,t), U_3(0,t))'$. The joint distribution of $\boldsymbol{U}(0,t)$ now becomes the key to obtain joint safety margins related to the different product types. With i.i.d. present values $\boldsymbol{U}^1(0,t), \boldsymbol{U}^2(0,t), \ldots$, the multivariate CLT now applies

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{U}^i(0,t) - \mathbb{E}\big[\boldsymbol{U}^1(0,t)\big]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\big(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}^1(0,t)}\big), \quad \text{as} \quad N \to \infty,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{U}^1(0,t)} = \big\{\mathrm{Cov}\big(U_i^1(0,t), U_j^1(0,t)\big)\big\}_{i,j=1,2,3}.$$

Hence, the covariance matrix of $\boldsymbol{U}^1(0,t)$ allows for approximation of joint safety margins for the three product types. Again, turning to higher order moments of $\boldsymbol{U}^1(0,t)$ in general improves these approximations in the same way as for the univariate case. The focal point of this paper is to compute these higher order moments for multivariate present values in a general multi-state framework, and give special attention to pair-wise covariances as motivated by this example.

We should like to mention that while this example concerned joint safety margins on the unsystematic biometric risk in life and disability insurance, one could have used the same model for e.g. systematic mortality risk. This is obtained by thinking of the states as demographic states, say, where e.g. the disability state represents a situation where the overall mortality for all insured in the portfolio have increased (or decreased). Then we would obtain joint safety margins related to systematic mortality risk when considering the joint distribution of the multivariate present value (a modification of the payment processes might be necessary here).

6.3 Setup

We now proceed with the general setup of the paper. Subsection 6.3.1 introduces the process governing the state of the insured and the multivariate payment process describing the collection of products held by the insured. Subsection 6.3.2 then considers the multivariate present value of future payments, whose components consists of the present value of the individual payments, and we introduce the calculation of their higher order moments as the main purpose of the paper. By considering existing results in Norberg (1995b) and Bladt, Asmussen, and Steffensen (2020) for single payment processes, we end the section by explaining our main ideas and give an overview on how we are to obtain the multivariate extension in this paper.

Notation and conventions For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we say that $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_\ell \leq y_\ell$ for all $\ell = 1, \ldots, n$. Furthermore, we denote with $\bar{\boldsymbol{x}}$ the sum of the elements in $\boldsymbol{x}$, i.e.

$\bar{\boldsymbol{x}} = \sum_{\ell=1}^{n} x_\ell$. Also, $\boldsymbol{e}_\ell$ is the ℓ'th unit vector in $\mathbb{R}^n$ and $E_n = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ denotes the set of unit vectors, i.e. the natural basis of $\mathbb{R}^n$. Furthermore, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{\ell=1}^{n} x_\ell y_\ell$ denotes the usual inner product.

For $n \times n$ matrix valued functions $\mathbb{R} \ni x \mapsto \boldsymbol{A}(x)$ we denote its product integral in the (time) interval $(s, t]$ as

$$\boldsymbol{F}(s,t) = \prod_{s}^{t} (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x),$$

whenever it exists; here $\boldsymbol{I}$ is the $n \times n$ identity matrix. For a survey on product integration and its properties, we refer to Gill and Johansen (1990) and Johansen (1986) and the applications to life insurance in Milbrodt and Stracke (1997) and Bladt, Asmussen, and Steffensen (2020). Numerical schemes for calculation of product integrals can be found in Helton and Stuckwisch (1976).

For the purpose of computing multivariate higher order moments, we consider the map $S : \mathbb{N}_0^n \to \mathcal{P}(\mathbb{N}_0^n)$ defined as

$$S(\boldsymbol{x}) = \left\{ \boldsymbol{\xi} \in \mathbb{N}_0^n \setminus \{\boldsymbol{0}\} \;\middle|\; \boldsymbol{\xi} \leq \boldsymbol{x} \right\}, \quad \boldsymbol{x} \in \mathbb{N}_0^n. \tag{6.3.1}$$

The set $S(\boldsymbol{x})$ then contains all combinations of lower order moments of $\boldsymbol{x}$, excluding the $\boldsymbol{0}$'th moment, and the cardinality $\left|S(\boldsymbol{x})\right| = \prod_{\ell=1}^{n}(x_\ell + 1) - 1$ then represents the amount of these that appear; whenever we wish to include the $\boldsymbol{0}$'th moment, we write $\widetilde{S}(\boldsymbol{k}) = S(\boldsymbol{k}) \cup \{\boldsymbol{0}\}$. For notational convenience, we simply define the mapping $|\cdot| : \mathbb{N}_0^n \to \mathbb{N}_0$ as

$$|\boldsymbol{x}| := \prod_{\ell=1}^{n}(x_\ell + 1) - 1, \quad \boldsymbol{x} \in \mathbb{N}_0^n.$$

To compute higher order partial derivatives of multivariate scalar functions $f : \mathbb{R}^n \to \mathbb{R}$, we use the following notation, for $\boldsymbol{m} \in \mathbb{N}_0^n$,

$$\frac{\partial^{\boldsymbol{m}}}{\partial \boldsymbol{x}^{\boldsymbol{m}}} f(\boldsymbol{x}) = \frac{\partial^{\bar{\boldsymbol{m}}}}{\partial x_1^{m_1} \partial x_2^{m_2} \cdot \ldots \cdot \partial x_n^{m_n}} f(\boldsymbol{x})$$

provided they exist.

6.3.1 Life insurance model

The state of the insured is governed by a non-explosive Markov jump process $Z = \{Z(t)\}_{t \geq 0}$ taking values on a finite state space $\mathcal{J} = \{0, \ldots, J-1\}$, $J \in \mathbb{N}$, with deterministic initial state $Z(0) = z_0 \in \mathcal{J}$. Denote with $\boldsymbol{M}(t) = \{\mu_{ij}(t)\}_{i,j \in \mathcal{J}}$ the transition intensity matrix of Z and let $\boldsymbol{P}(s,t) = \{p_{ij}(s,t)\}_{i,j \in \mathcal{J}}$ be the corresponding transition probability matrix. The multivariate counting process associated with Z is denoted $\boldsymbol{N}(t) = \{N_{ij}(t)\}_{i,j \in \mathcal{J}, i \neq j}$, with components given by

$$N_{ij}(t) = \#\left\{ s \in (0,t] : Z(s-) = i, Z(s) = j \right\}.$$

We consider a collection of life insurance contracts described by the n-dimensional payment process $\boldsymbol{B}(t) = (B_1(t), \dots, B_n(t))'$ with each component giving accumulated benefits less premiums for the corresponding contract; here $n \in \mathbb{N}$ is a fixed and finite number of contracts. We suppose that the ℓ'th payment process, $\ell = 1, \dots, n$, consist of deterministic sojourn payment rates b_i^ℓ and transition payments b_{ij}^ℓ, $i, j \in \mathcal{J}, i \neq j$, in terms of the state process Z. Coupling these in vectors,

$$\boldsymbol{b}_i(t) = \left(b_i^1(t), \dots, b_i^n(t)\right)', \tag{6.3.2}$$

$$\boldsymbol{b}_{ij}(t) = (b_{ij}^1(t), \dots, b_{ij}^n(t))', \tag{6.3.3}$$

we formally characterize $\boldsymbol{B}$ as

$$\mathrm{d}\boldsymbol{B}(t) = \sum_{i \in \mathcal{J}} \left(\mathbb{1}_{(Z(t)=i)} \boldsymbol{b}_i(t) \mathrm{d}t + \sum_{j: j \neq i} \boldsymbol{b}_{ij}(t) \mathrm{d}N_{ij}(t) \right), \quad \boldsymbol{B}(0) = \mathbf{0}. \tag{6.3.4}$$

We assume throughout that the transition rates μ_{ij} as well as the payment functions b_i^ℓ, b_{ij}^ℓ, for $i, j \in \mathcal{J}$, $i \neq j$, and $\ell = 1, \dots, n$, are bounded on finite intervals. This makes all integrals and expectations in the following well defined and finite.

We think of $\boldsymbol{B}$ as describing payments with different product types that naturally are dependent. The simplest example is the combination of life annuities and death benefit, which have opposite effects with respect to the mortality risk on lifetimes. In the general multi-state framework of this paper, we are able to examine product interactions that are less trivial; cf. the motivating example in Section 6.2.

Example 6.3.1 (Disability model)**.** The setup of the motivating example from Section 6.2 in the disability model corresponds to having the following vectors of payments functions (6.3.2)–(6.3.3):

$$\begin{aligned} \boldsymbol{b}_0(t) &= \left(0, b \cdot \mathbb{1}_{(t \geq T)}, 0\right)', \\ \boldsymbol{b}_1(t) &= \left(0, b \cdot \mathbb{1}_{(t \geq T)}, d \cdot \mathbb{1}_{(t < T)}\right)', \\ \boldsymbol{b}_{02}(t) = \boldsymbol{b}_{12}(t) &= \left(S \cdot \mathbb{1}_{(t < T)}, 0, 0\right)', \end{aligned}$$

with the remaining vectors of payment functions having purely zero entries. ○

6.3.2 Present values and their moments

Let $r = \{r(t)\}_{t \geq 0}$ be a continuous, suitably regular and deterministic short rate. To account for time value of money, introduce $v(s,t)$ as the discount factor for a payment at time t valuated at time $s \leq t$, i.e.

$$v(s,t) = e^{-\int_s^t r(x)\,\mathrm{d}x}.$$

The present value at time s of future benefits less premiums up to time t for the payment process $\boldsymbol{B}$ is then given by

$$\boldsymbol{U}(s,t) = \int_s^t v(s,u)\,\mathrm{d}\boldsymbol{B}(u) = (U_1(s,t), \dots, U_n(s,t))', \tag{6.3.5}$$

where the components $U_\ell(s,t) = \int_s^t v(s,u)\,\mathrm{d}B_\ell(u)$, $\ell = 1,\ldots,n$, are the present values of the individual payment process.

The components of $\boldsymbol{U}$ are (by construction) seen to be dependent, which introduces joint effects on the payments with respect to the underlying event risk. To analyze these dependence structures, we wish to examine its joint distribution by computing its conditional higher order moments given the current state of the insured, that is, by computing

$$V_i^{(\boldsymbol{k})}(s,t) = \mathbb{E}\left[\prod_{\ell=1}^{n} U_\ell(s,t)^{k_\ell} \,\middle|\, Z(s) = i\right] \tag{6.3.6}$$

for any $i \in \mathcal{J}$ and $\boldsymbol{k} = (k_1,\ldots,k_n) \in \mathbb{N}_0^n$. For $\boldsymbol{k} = \boldsymbol{e}_\ell$ we have that $V_i^{(\boldsymbol{e}_\ell)}(s,t)$ is the state-wise prospective reserve for the ℓ'th payment process, calculated using the classic Thiele's differential equation, cf. Hoem (1969a).

As mentioned in the introduction, having multiple product types for the same insured is usually handled by adding all payment functions together as $\bar{\boldsymbol{b}}_i$ and $\bar{\boldsymbol{b}}_{ij}$, or, equivalently, by considering the aggregate present value $\bar{\boldsymbol{U}}(s,t)$. This effectively gives a one dimensional payment process to examine; in Hoem (1969a) we have Thiele's differential equation for conditional expectations and in Adékambi and Christiansen (2017), Bladt, Asmussen, and Steffensen (2020), and Norberg (1995b) we have differential equations for conditional higher order moments, which e.g. allows for calculation of safety margins on reserves as in Christiansen (2013). The present paper is therefore to be thought of as a way of decomposing the aggregated payments into the different product types and study the joint distribution of their future payments, which then would allow for calculation of joint safety margins for the different product types as outlined in Section 6.2.

The following matrices, which we think of as inputs to the computations, are needed in the derivations; for $\ell \in \{1,\ldots,n\}$, $i,j \in \mathcal{J}, i \neq j$, and $\boldsymbol{y} \in \mathbb{N}_0^n \setminus (E_n \cup \{\boldsymbol{0}\})$, let

$$\boldsymbol{B}_\ell(s) = \left\{b_{ij}^\ell(s)\right\}_{i,j\in\mathcal{J}}, \tag{6.3.7}$$

$$\boldsymbol{b}^\ell(s) = \left(b_0^\ell(s),\ldots,b_{J+1}^\ell(s)\right)', \tag{6.3.8}$$

$$\boldsymbol{R}_\ell(s) = \boldsymbol{M}(s) \bullet \boldsymbol{B}_\ell(s) + \boldsymbol{\Delta}(\boldsymbol{b}^\ell(s)), \tag{6.3.9}$$

$$\boldsymbol{C}^{(\boldsymbol{y})}(s) = \boldsymbol{M}(s) \bullet \boldsymbol{B}_1^{\bullet y_1}(s) \bullet \ldots \bullet \boldsymbol{B}_n^{\bullet y_n}(s), \tag{6.3.10}$$

where $\Delta(\boldsymbol{b}^\ell(s))$ denotes the diagonal matrix with the vector $\boldsymbol{b}^\ell(s)$ as diagonal elements, $\bullet$ denotes the entry-wise matrix product and e.g. $\boldsymbol{B}_1^{\bullet y}(s)$ denotes the entry-wise matrix product of $\boldsymbol{B}_1(s)$ with itself y times; here we use the convention that e.g. $\boldsymbol{B}_1^{\bullet 0}(s)$ denotes the matrix with all entries equal to 1. These matrices are defined analogously to those of Bladt, Asmussen, and Steffensen (2020, (3.8)–(3.11)) with the relevant extension to a multivariate payment process.

The remainder of the paper now focuses on computation of $\boldsymbol{V}_i^{(\boldsymbol{k})}(s,t)$, carried out as follows. Following Bladt, Asmussen, and Steffensen (2020), we first compute the conditional higher order moments on events that the insured terminates in specific states, given by, for $j \in \mathcal{J}$,

$$V_{ij}^{(\boldsymbol{k})}(s,t) = \mathbb{E}\left[\mathbb{1}_{(Z(t)=j)} \prod_{\ell=1}^{n} U_\ell(s,t)^{k_\ell} \,\middle|\, Z(s) = i \right] \tag{6.3.11}$$

with corresponding matrix

$$\boldsymbol{V}^{(\boldsymbol{k})}(s,t) = \left\{ V_{ij}^{(\boldsymbol{k})}(s,t) \right\}_{i,j\in\mathcal{J}}. \tag{6.3.12}$$

This is carried out by deriving the relevant moment generating function. As in Bladt, Asmussen, and Steffensen (2020), we shall refer to these as the *conditional partial moments* of $\boldsymbol{U}(s,t)$. From these, the conditional moments $V_i^{(\boldsymbol{k})}(s,t)$ are obtained via the relation $\sum_{j\in\mathcal{J}} V_{ij}^{(\boldsymbol{k})}(s,t) = V_i^{(\boldsymbol{k})}(s,t)$, such that

$$V_i^{(\boldsymbol{k})}(s,t) = \left\{ \boldsymbol{V}^{(\boldsymbol{k})}(s,t)\mathbf{1} \right\}_i, \tag{6.3.13}$$

where $\mathbf{1} = (1,1,\dots,1)'$.

For the first order moment $k = 1$ (in the one-dimensional case $n = 1$), the partial reserves (6.3.11) have already been introduced in existing literature; they coincide with the conditional prospective premium reserves introduced by Wolthuis (1992). In Norberg (1991), we also encounter these type of quantities for corresponding retrospective reserves initiated at time zero, i.e. for $s = 0$.

These earlier treatments in the literature of these kinds of quantities justifies the conditional partial moments (6.3.12) as relevant objects to study on their own. In this paper, however, they are primarily introduced as a mathematical convenient object towards obtaining the conditional moments (6.3.6); this is the same focus in Bladt, Asmussen, and Steffensen (2020).

6.4 Moment generating functions

Consider the multivariate present value $\boldsymbol{U}(s,t)$. The moment generating function of its conditional distribution given $Z(s) = i$ on the event $(Z(t) = j)$, $i,j \in \mathcal{J}$, is given by

$$F_{ij}(\boldsymbol{\theta};s,t) = \mathbb{E}\left[e^{\langle \boldsymbol{\theta}, \boldsymbol{U}(s,t)\rangle} \mathbb{1}_{(Z(t)=j)} \,\middle|\, Z(s) = i \right], \quad \boldsymbol{\theta} \in \mathbb{R}^n, \tag{6.4.1}$$

with corresponding matrix

$$\boldsymbol{F}(\boldsymbol{\theta};s,t) = \left\{ F_{ij}(\boldsymbol{\theta};s,t) \right\}_{i,j\in\mathcal{J}}.$$

We are then able to obtain the moment generating function by a direct application of results in Bladt, Asmussen, and Steffensen (2020).

Theorem 6.4.1. *The joint distribution of the multivariate present value $\boldsymbol{U}(s,t)$ has moment generating function given by*

$$\boldsymbol{F}(\boldsymbol{\theta};s,t) = \prod_s^t \left(\boldsymbol{I} + \left[\boldsymbol{M}(u) \bullet \left\{e^{v(s,u)\langle \boldsymbol{\theta}, \boldsymbol{b}_{ij}(u)\rangle}\right\}_{i,j\in\mathcal{J}} + v(s,u)\sum_{\ell=1}^n \theta_\ell \boldsymbol{\Delta}(\boldsymbol{b}^\ell(u))\right] \mathrm{d}u\right).$$

Proof. Let $\boldsymbol{\theta} \in \mathbb{R}^n$ and $0 \leq s \leq t$ be given. By linearity of integrals, we note

$$\begin{aligned}\langle \boldsymbol{\theta}, \boldsymbol{U}(s,t)\rangle &= \int_s^t v(s,u)\,\mathrm{d}\widetilde{B}_{\boldsymbol{\theta}}(u),\\ \mathrm{d}\widetilde{B}_{\boldsymbol{\theta}}(u) &= \sum_{i\in\mathcal{J}} 1_{\{Z(u-)=i\}}\left(\langle \boldsymbol{\theta}, \boldsymbol{b}_i(u)\rangle \mathrm{d}u + \sum_{j:j\neq i}\langle \boldsymbol{\theta}, \boldsymbol{b}_{ij}(u)\rangle \mathrm{d}N_{ij}(u)\right).\end{aligned}$$

Hence, $\langle \boldsymbol{\theta}, \boldsymbol{U}(s,t)\rangle$ is the present value of a one-dimensional payment process with sojourn payment rates $\langle \boldsymbol{\theta}, \boldsymbol{b}_i(\cdot)\rangle$ and transition payments $\langle \boldsymbol{\theta}, \boldsymbol{b}_{ij}(\cdot)\rangle$. Thus, in the spirit of Bladt, Asmussen, and Steffensen (2020), we may view $\boldsymbol{F}(\boldsymbol{\theta};s,t)$ as the moment generating function of a total (undiscounted) reward evaluated in 1 with the payment functions $u \mapsto \langle \boldsymbol{\theta}, v(s,u)\boldsymbol{b}_i(u)\rangle$ and $u \mapsto \langle \boldsymbol{\theta}, v(s,u)\boldsymbol{b}_{ij}(u)\rangle$ (for fixed s). The result then follows from an application of Bladt, Asmussen, and Steffensen (2020, Theorem 3). □

Remark 6.4.2. Theorem 6.4.1 is a generalization of Bladt, Asmussen, and Steffensen (2020, Theorem 3) to multivariate payment processes with discounting. Indeed, by letting $n = 1$ and $r(s) = 0$ for all $s \geq 0$, the result simplifies to, suppressing the superfluous dependency in ℓ,

$$\boldsymbol{F}(\theta;s,t) = \prod_s^t \left(\boldsymbol{I} + \left[\boldsymbol{M}(s) \bullet \left\{e^{\theta b_{ij}(s)}\right\}_{i,j\in\mathcal{J}} + \theta\boldsymbol{\Delta}(\boldsymbol{b}(s))\right]\mathrm{d}u\right),$$

which is the product integral of Bladt, Asmussen, and Steffensen (2020, Theorem 3). △

From the product integral representation of $\boldsymbol{F}$, we are able to obtain a partial differential equation satisfied by $\boldsymbol{F}$. Similar type of equations are also obtained in Adékambi and Christiansen (2017).

Theorem 6.4.3. *The moment generating function $\boldsymbol{F}$ satisfies the system of partial differential equations*

$$\begin{aligned}\frac{\partial}{\partial s}\boldsymbol{F}(\boldsymbol{\theta};s,t) &= r(s)\sum_{\ell=1}^n \theta_\ell \frac{\partial}{\partial \theta_\ell}\boldsymbol{F}(\boldsymbol{\theta};s,t)\\ &\quad - \left[\boldsymbol{M}(s) \bullet \left\{e^{\langle \boldsymbol{\theta}, \boldsymbol{b}_{ij}(s)\rangle}\right\}_{i,j\in\mathcal{J}} + \sum_{\ell=1}^n \theta_\ell \boldsymbol{\Delta}(\boldsymbol{b}^\ell(s))\right]\boldsymbol{F}(\boldsymbol{\theta};s,t),\end{aligned} \tag{6.4.2}$$

with the terminal condition $\boldsymbol{F}(\boldsymbol{\theta};t,t) = \boldsymbol{I}$.

Proof. For $\boldsymbol{\theta} \in \mathbb{R}^n$ and $0 \leq s \leq u$, let

$$\boldsymbol{A}(s,u;\boldsymbol{\theta}) = \boldsymbol{M}(u) \bullet \left\{ e^{v(s,u)\langle \boldsymbol{\theta}, \boldsymbol{b}_{ij}(u)\rangle} \right\}_{i,j\in\mathcal{J}} + v(s,u)\sum_{\ell=1}^{n} \theta_\ell \boldsymbol{\Delta}(\boldsymbol{b}^\ell(u))$$

denote the matrix that is product integrated in Theorem 6.4.1. It satisfies the relation $\frac{\partial}{\partial s}\boldsymbol{A}(s,u;\boldsymbol{\theta}) = r(s)\sum_{\ell=1}^n \theta_\ell \frac{\partial}{\partial \theta_\ell}\boldsymbol{A}(s,u;\boldsymbol{\theta})$, which can be seen from direct calculations. Now, since $\boldsymbol{F}$ is a product integral of the matrix function $\boldsymbol{A}(s,\cdot;\boldsymbol{\theta})$, it has the Peano-Baker series representation given by

$$\boldsymbol{F}(\boldsymbol{\theta};s,t) = \boldsymbol{I} + \sum_{n=1}^{\infty}\int_s^t \int_s^{x_n} \cdots \int_s^{x_2} \boldsymbol{A}(s,x_1;\boldsymbol{\theta})\cdots\boldsymbol{A}(s,x_n;\boldsymbol{\theta})\,\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

Differentiating this w.r.t. s, we get using Lebniz' integral rule,

$$\begin{aligned}\frac{\partial}{\partial s}\boldsymbol{F}(\boldsymbol{\theta};s,t) &= -\boldsymbol{A}(s,s;\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta};s,t)\\ &\quad + \sum_{n=1}^{\infty}\int_s^t \int_s^{x_n} \cdots \int_s^{x_2} \frac{\partial}{\partial s}\left(\boldsymbol{A}(s,x_1;\boldsymbol{\theta})\cdots\boldsymbol{A}(s,x_n;\boldsymbol{\theta})\right)\mathrm{d}x_1\cdots\mathrm{d}x_n\end{aligned}$$

In the last differentiation, we can in each combination (when applying the product rule) substitute the derivative w.r.t. to s with $r(s)\sum_{\ell=1}^n \theta_\ell \frac{\partial}{\partial \theta_\ell}\boldsymbol{A}(s,u;\boldsymbol{\theta})$. This gives

$$\begin{aligned}&\frac{\partial}{\partial s}\boldsymbol{F}(\boldsymbol{\theta};s,t)\\ &= -\boldsymbol{A}(s,s;\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta};s,t)\\ &+ \sum_{n=1}^{\infty}\int_s^t \int_s^{x_n} \cdots \int_s^{x_2} r(s)\sum_{\ell=1}^{n}\theta_\ell\frac{\partial}{\partial \theta_\ell}\left(\boldsymbol{A}(s,x_1;\boldsymbol{\theta})\cdots\boldsymbol{A}(s,x_n;\boldsymbol{\theta})\right)\mathrm{d}x_1\cdots\mathrm{d}x_n\\ &= -\boldsymbol{A}(s,s;\boldsymbol{\theta})\boldsymbol{F}(\boldsymbol{\theta};s,t) + r(s)\sum_{\ell=1}^{n}\theta_\ell\frac{\partial}{\partial \theta_\ell}\boldsymbol{F}(\boldsymbol{\theta};s,t),\end{aligned}$$

as claimed. The boundary condition follows from the definition of product integrals. Throughout the proof, we have used that the Peano-Baker series converges uniformly on compact intervals in order to interchange summation and differentiation. □

Remark 6.4.4. Consider the univariate case $n = 1$ and suppress the superfluous dependency in ℓ. By multiplying both sides of (6.4.2) with $\mathbf{1} = (1,1,\ldots,1)'$ and extracting the i'th element, $i \in \mathcal{J}$, we see that $F_i(\theta;s,t) := \mathbb{E}\left[e^{\theta U(s,t)} \,\middle|\, Z(s) = i\right]$ satisfies the partial differential equation

$$\begin{aligned}\frac{\partial}{\partial s}F_i(\theta;s,t) &= r(s)\theta\frac{\partial}{\partial \theta}F_i(\theta;s,t) - \theta b_i(s)F_i(\theta;s,t)\\ &\quad - \sum_{k:k\neq i}\mu_{ik}(s)\left(e^{\theta b_{ik}(s)}F_k(\theta;s,t) - F_i(\theta;s,t)\right),\end{aligned}$$

with the boundary condition $F_i(\theta; t, t) = 1$. This is the partial differential equation of Adékambi and Christiansen (2017, Proposition 4.3) in the Markovian special case with no duration dependency. △

Having derived the relevant moment generating function of the multivariate present value $\boldsymbol{U}(s,t)$, we are now ready to compute the corresponding higher order moments.

6.5 Higher order moments

In this section, we derive the conditional (partial) higher order moments (6.3.6) and (6.3.12), which constitutes the main contribution of the paper. In Subsection 6.5.1 we derive differential equations for the conditional partial moments (6.3.12) using the moment generating function $\boldsymbol{F}(\cdot; s, t)$ derived in Section 6.4. Then we use this result to derive differential equations for the conditional moments (6.3.6) via the relation (6.3.13). In Subsection 6.5.2, we then show how all conditional partial moments up to a given order $\boldsymbol{k} \in \mathbb{N}_0^n$ can be obtained from a product integral calculation. Throughout the section, we highlight how the results appear as natural and simple multivariate extensions to results in Bladt, Asmussen, and Steffensen (2020) and Norberg (1995b), in particular when notation may appear as cumbersome.

6.5.1 Differential equations of conditional (partial) moments

We now turn our attention to deriving ordinary differential equations for the conditional (partial) moments. For this, recall that $S(\boldsymbol{k})$ denotes the set of all lower order moments of the $\boldsymbol{k}$'th moment, excluding the $\mathbf{0}$'th moment, with elements we may write as

$$S(\boldsymbol{k}) = \left\{ \boldsymbol{y}^1, \ldots, \boldsymbol{y}^{|\boldsymbol{k}|} \right\}.$$

The following main result then take use of the moment generating function derived in Section 6.4 to obtain differential equations for the conditional partial moments.

Theorem 6.5.1. *The conditional partial moments $\boldsymbol{V}^{(\boldsymbol{k})}(\cdot, t)$ satisfies the backward differential equations given by, for $\boldsymbol{k} \in \mathbb{N}_0^n$,*

$$\begin{aligned} \frac{\partial}{\partial s} \boldsymbol{V}^{(\boldsymbol{k})}(s,t) = {} & \big(\bar{\boldsymbol{k}}\, r(s)\boldsymbol{I} - \boldsymbol{M}(s)\big)\boldsymbol{V}^{(\boldsymbol{k})}(s,t) - \sum_{\ell=1}^{n} k_\ell \boldsymbol{R}_\ell(s) \boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{e}_\ell)}(s,t) \\ & - \sum_{\substack{\boldsymbol{y}\in S(\boldsymbol{k}) \\ \boldsymbol{y}\notin E_n}} \prod_{\ell=1}^{n} \binom{k_\ell}{y_\ell} \boldsymbol{C}^{(\boldsymbol{y})}(s) \boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(s,t), \end{aligned} \tag{6.5.1}$$

with the terminal conditions $\boldsymbol{V}^{(\boldsymbol{k})}(t,t) = \mathbb{1}_{(\boldsymbol{k}=\mathbf{0})}\boldsymbol{I}$.

Proof. We can use the moment generating function $\boldsymbol{F}(\cdot\,; s,t)$ to calculate derivative $\frac{\partial}{\partial s}\boldsymbol{V}^{(\boldsymbol{k})}(s,t)$ as

$$\frac{\partial}{\partial s}\boldsymbol{V}^{(\boldsymbol{k})}(s,t) = \frac{\partial^{\boldsymbol{k}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}}}\frac{\partial}{\partial s}\boldsymbol{F}(\boldsymbol{\theta}; s,t)\bigg|_{\boldsymbol{\theta}=\boldsymbol{0}}.$$

Differentiating both sides of (6.4.2), we get by the generalized product rule

$$\begin{aligned}
&\frac{\partial^{\boldsymbol{k}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}}}\frac{\partial}{\partial s}\boldsymbol{F}(\boldsymbol{\theta}; s,t)\\
&= r(s)\sum_{\ell=1}^{n}\sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})}\prod_{r=1}^{n}\binom{k_r}{y_r}\left[\frac{\partial^{\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{y}}}\theta_\ell\right]\frac{\partial^{\boldsymbol{k}-\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}-\boldsymbol{y}}}\frac{\partial}{\partial\theta_\ell}\boldsymbol{F}(\boldsymbol{\theta}; s,t)\\
&- \sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})}\prod_{r=1}^{n}\binom{k_r}{y_r}\frac{\partial^{\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{y}}}\left[\boldsymbol{M}(s)\bullet\left\{e^{\langle\boldsymbol{\theta},\boldsymbol{b}_{ij}(s)\rangle}\right\}_{i,j\in\mathcal{J}} + \sum_{\ell=1}^{n}\theta_\ell\boldsymbol{\Delta}(\boldsymbol{b}^\ell(s))\right]\times\\
&\qquad\qquad\frac{\partial^{\boldsymbol{k}-\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}-\boldsymbol{y}}}\boldsymbol{F}(\boldsymbol{\theta}; s,t).\\
&= r(s)\sum_{\ell=1}^{n}\left[\mathbb{1}_{(\boldsymbol{y}=\boldsymbol{0})}\theta_\ell + k_\ell\mathbb{1}_{(\boldsymbol{y}=\boldsymbol{e}_\ell)}\right]\frac{\partial^{\boldsymbol{k}-\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}-\boldsymbol{y}}}\frac{\partial}{\partial\theta_\ell}\boldsymbol{F}(\boldsymbol{\theta}; s,t)\\
&- \sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})}\prod_{r=1}^{n}\binom{k_r}{y_r}\Bigg[\boldsymbol{M}(s)\bullet\left\{\prod_{\ell=1}^{n}b_{ij}^\ell(s)^{y_\ell}e^{\langle\boldsymbol{\theta},\boldsymbol{b}_{ij}(s)\rangle}\right\}_{i,j\in\mathcal{J}}\\
&\qquad\qquad + \sum_{\ell=1}^{n}\mathbb{1}_{(\boldsymbol{y}=\boldsymbol{e}_\ell)}\boldsymbol{\Delta}(\boldsymbol{b}^\ell(s))\Bigg]\frac{\partial^{\boldsymbol{k}-\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}-\boldsymbol{y}}}\boldsymbol{F}(\boldsymbol{\theta}; s,t).
\end{aligned}$$

In the last equality, we have used that $\prod_{r=1}^{n}\binom{k_r}{y_r} = k_\ell$ if $\boldsymbol{y} = \boldsymbol{e}_\ell$ along with $\frac{\partial^{\boldsymbol{y}}}{\partial\boldsymbol{\theta}^{\boldsymbol{y}}}\theta_\ell = \mathbb{1}_{(\boldsymbol{y}=\boldsymbol{0})}\theta_\ell + \mathbb{1}_{(\boldsymbol{y}=\boldsymbol{e}_\ell)}$ for $\ell = 1,\ldots,n$. Now, evaluating in $\boldsymbol{\theta} = \boldsymbol{0}$, we get

$$\begin{aligned}
&\frac{\partial}{\partial s}\boldsymbol{V}^{(\boldsymbol{k})}(s,t)\\
&= r(s)\sum_{\ell=1}^{n}k_\ell\frac{\partial^{\boldsymbol{k}-\boldsymbol{e}_\ell}}{\partial\boldsymbol{\theta}^{\boldsymbol{k}-\boldsymbol{e}_\ell}}\frac{\partial}{\partial\theta_\ell}\boldsymbol{F}(\boldsymbol{\theta}; s,t)\bigg|_{\boldsymbol{\theta}=\boldsymbol{0}}\\
&\quad -\sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})}\prod_{\ell=1}^{n}\binom{k_\ell}{y_\ell}\left[\boldsymbol{M}(s)\bullet\left\{\prod_{\ell=1}^{n}b_{ij}^\ell(s)^{y_\ell}\right\}_{i,j\in\mathcal{J}} + \sum_{\ell=1}^{n}\mathbb{1}_{(\boldsymbol{y}=\boldsymbol{e}_\ell)}\boldsymbol{\Delta}(\boldsymbol{b}^\ell(s))\right]\times\\
&\qquad\qquad\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(s,t)\\
&= \bar{\boldsymbol{k}}\,r(s)\boldsymbol{V}^{(\boldsymbol{k})}(s,t)\\
&\quad -\sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})}\prod_{\ell=1}^{n}\binom{k_\ell}{y_\ell}\left[\boldsymbol{M}(s)\bullet B_1^{\bullet y_1}(s)\bullet\ldots\bullet B_n^{\bullet y_n}(s) + \sum_{\ell=1}^{n}\mathbb{1}_{(\boldsymbol{y}=\boldsymbol{e}_\ell)}\boldsymbol{\Delta}(\boldsymbol{b}^\ell(s))\right]\times\\
&\qquad\qquad\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(s,t).
\end{aligned}$$

Splitting the last sum into terms corresponding to $\boldsymbol{y} = \boldsymbol{0}$ and $\boldsymbol{y} = \boldsymbol{e}_\ell$, for $\ell = 1, \dots, n$, gives

$$\frac{\partial}{\partial s}\boldsymbol{V}^{(\boldsymbol{k})}(s,t) = \big(\bar{\boldsymbol{k}}\, r(s)\boldsymbol{I} - \boldsymbol{M}(s)\big)\boldsymbol{V}^{(\boldsymbol{k})}(s,t) - \sum_{\ell=1}^{n} k_\ell \boldsymbol{R}_\ell(s)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{e}_\ell)}(s,t)$$
$$- \sum_{\substack{\boldsymbol{y}\in S(\boldsymbol{k})\\ \boldsymbol{y}\notin E_n}} \prod_{\ell=1}^{n} \binom{k_\ell}{y_\ell} \boldsymbol{C}^{(\boldsymbol{y})}(s)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(s,t),$$

the desired differential equation. The boundary condition follows from (6.3.11) with the fact that $\boldsymbol{V}^{(\boldsymbol{0})}(t,t) = \boldsymbol{P}(t,t) = \boldsymbol{I}$. □

Remark 6.5.2. The differential equation of Theorem 6.5.1 generalizes Norberg (1995b, Theorem 6) to multivariate present values; for $n = 1$ we have, supressing the superfluous dependency in ℓ, that $S(k) = \{1, 2, \dots, k\}$ and $E_1 = \{1\}$, and so the differential equations reads

$$\frac{\partial}{\partial s}\boldsymbol{V}^{(k)}(s,t) = (k\, r(s)\boldsymbol{I} - \boldsymbol{M}(s))\boldsymbol{V}^{(k)}(s,t) - k\boldsymbol{R}(s)\boldsymbol{V}^{(k-1)}(s,t)$$
$$- \sum_{y=2}^{k} \binom{k}{y} \boldsymbol{C}^{(y)}(s)\boldsymbol{V}^{(k-y)}(s,t),$$

with the terminal conditions $\boldsymbol{V}^{(k)}(t,t) = \mathbb{1}_{(k=0)}\boldsymbol{I}$. △

From the differential equation of the conditional partial moments, we immediately obtain differential equations for the corresponding conditional moments $V_i^{(\boldsymbol{k})}(s,t)$ by multiplying the former with $\boldsymbol{1} = (1, 1, \dots, 1)'$, as explained in Section 6.3.

Theorem 6.5.3. *The conditional moments $V_i^{(\boldsymbol{k})}(\cdot, t)$ satisfies the backward differential equations given by, for $i \in \mathcal{J}$,*

$$\frac{\partial}{\partial s}V_i^{(\boldsymbol{k})}(s,t) = \big(\bar{\boldsymbol{k}} r(s) + \mu_{i\cdot}(s)\big)V_i^{(\boldsymbol{k})}(s,t) - \sum_{\ell=1}^{n} k_\ell\, b_i^\ell(s) V_i^{(\boldsymbol{k}-\boldsymbol{e}_\ell)}(s,t)$$
$$- \sum_{j:j\neq i} \mu_{ij}(s) \sum_{\boldsymbol{y}\in\widetilde{S}(\boldsymbol{k})} \prod_{\ell=1}^{n} \binom{k_\ell}{y_\ell} (b_{ij}^\ell(s))^{y_\ell} V_j^{(\boldsymbol{k}-\boldsymbol{y})}(s,t), \tag{6.5.2}$$

with the terminal conditions $V_i^{(\boldsymbol{k})}(t,t) = \mathbb{1}_{(\boldsymbol{k}=\boldsymbol{0})}$.

Proof. Simply multiply the differential equation in Theorem 6.5.1 with the vector $\boldsymbol{1} = (1, 1, \dots, 1)'$ on both sides and extract the i'th element of the vectors. □

Remark 6.5.4. The differential equation of Theorem 6.5.3 generalizes Norberg (1995b, (3.2)) to multivariate present values; for $n = 1$ we have, supressing the superfluous dependency in ℓ, that $S(k) = \{1, 2, \ldots, k\}$ and $E_1 = \{1\}$, and so the differential equations reads

$$\begin{aligned}\frac{\partial}{\partial s} V_i^{(k)}(s,t) &= (kr(s) + \mu_{i\cdot}(s))V_i^{(k)}(s,t) - kb_i(s)V_i^{(k-1)}(s,t) \\ &\quad - \sum_{j:j\neq i} \mu_{ij}(s) \sum_{y=0}^{k} \binom{k}{y} (b_{ij}(s))^y V_j^{(k-y)}(s,t),\end{aligned}$$

with the terminal conditions $V_i^{(k)}(t,t) = \mathbb{1}_{(k=0)}$. △

Remark 6.5.5. The conditional central moments

$$m_i^{(\boldsymbol{k})}(s,t) = \mathbb{E}\left[\prod_{\ell=1}^{n} \Big(U_\ell(s,t) - V_i^{(\boldsymbol{e}_\ell)}(s,t) \Big)^{k_\ell} \,\middle|\, Z(s) = i \right]$$

may be obtained from the conditional (non-central) moments $V_i^{(\boldsymbol{k})}(s,t)$ via an application of the multidimensional binomial formula, giving

$$m_i^{(\boldsymbol{k})}(s,t) = \sum_{\boldsymbol{y} \in \widetilde{S}(\boldsymbol{k})} \prod_{\ell=1}^{n} (-1)^{k_\ell - y_\ell} \binom{k_\ell}{y_\ell} V_i^{(\boldsymbol{y})}(s,t) V_i^{(\boldsymbol{e}_\ell)}(s,t)^{k_\ell - y_\ell}. \tag{6.5.3}$$

Thus, by first solving the differential equation of Theorem 6.5.3 we are able to compute $m_i^{(\boldsymbol{k})}(s,t)$ via (6.5.3). Here, we may note that solving the differential equation of Theorem 6.5.3 to compute the $\boldsymbol{k}$'th moment immediately gives $V_i^{(\boldsymbol{y})}(s,t)$ for all $\boldsymbol{y} \in \widetilde{S}(\boldsymbol{k})$; details regarding this is explained in Subsection 6.5.2. △

Example 6.5.6 (Conditional covariance)**.** For $n = 2$ and $\boldsymbol{k} = (1,1)$ we have that $\widetilde{S}(\boldsymbol{k}) = \{(0,0), (1,0), (0,1), (1,1)\}$, and so the differential equation for the conditional product moment $V_i^{(1,1)}(s,t) = \mathbb{E}[U_1(s,t)U_2(s,t) \mid Z(s) = i]$ is given by

$$\begin{aligned}\frac{\partial}{\partial s} V_i^{(1,1)}(s,t) &= (2r(s) + \mu_{i\cdot}(s))V_i^{(1,1)}(s,t) - b_i^1(s)V_i^{(0,1)}(s,t) - b_i^2(s)V_i^{(1,0)}(s,t) \\ &\quad - \sum_{j:j\neq i} \mu_{ij}(s) \Big(V_j^{(1,1)}(s,t) + b_{ij}^1(s)V_j^{(0,1)}(s,t) + b_{ij}^2(s)V_j^{(1,0)}(s,t) \\ &\qquad + b_{ij}^1(s)b_{ij}^2(s) \Big), \\ V_i^{(1,1)}(t,t) &= 0,\end{aligned}$$

where $V_i^{(0,1)}$ and $V_i^{(1,0)}$ are the state-wise prospective reserves of the two payment processes, which can be calculated using Thiele's differential equation. From this, the conditional central moment $m_i^{(1,1)}(s,t)$ is given as

$$\begin{aligned}m_i^{(1,1)}(s,t) &= \mathbb{E}\Big[\Big(U_1(s,t) - V_i^{(1,0)}(s,t) \Big) \Big(U_2(s,t) - V_i^{(0,1)}(s,t) \Big) \,\Big|\, Z(s) = i \Big] \\ &= V_i^{(1,1)}(s,t) - V_i^{(1,0)}(s,t)V_i^{(0,1)}(s,t),\end{aligned} \tag{6.5.4}$$

which is the conditional covariance of $U_1(s,t)$ og $U_2(s,t)$ given $Z(s) = i$. ∘

Following up on the example, it turns out that it is possible to derive differential equations for the conditional covariance between two present values in which the sum at risk used in the Thiele differential equations appears. The result is also presented without proof in Asmussen and Steffensen (2020, Proposition 10.3), and we are able to proof the result here through our results on the non-central moments.

Corollary 6.5.7. *The conditional covariance $m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t)$ between $U_\ell(s,t)$ and $U_m(s,t)$ given $Z(s)=i$ satisfies the system of backward differential equations given by, for $i \in \mathcal{J}$,*

$$
\begin{aligned}
\frac{\partial}{\partial s} m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) &= 2r(s) m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) \\
&\quad - \sum_{j:j\neq i} \mu_{ij}(s)\Big(R_{ij}^\ell(s)R_{ij}^m(s) + m_j^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) - m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t)\Big),
\end{aligned}
$$

with terminal conditions $m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(t,t)=0$. Here, R_{ij}^ℓ denotes the sum at risk in the ℓ'th payment process for the transition from state i to j, $i\neq j$, given by

$$
R_{ij}^\ell(s) = b_{ij}^\ell(s) + V_j^{(\boldsymbol{e}_\ell)}(s,t) - V_i^{(\boldsymbol{e}_\ell)}(s,t),
$$

and analogously for R_{ij}^m.

Proof. Differentiating (6.5.4) w.r.t. s on both sides gives

$$
\begin{aligned}
\frac{\partial}{\partial s} m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) &= \frac{\partial}{\partial s} V_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) - V_i^{(\boldsymbol{e}_m)}(s,t)\frac{\partial}{\partial s}V_i^{(\boldsymbol{e}_\ell)}(s,t) \\
&\quad - V_i^{(\boldsymbol{e}_\ell)}(s,t)\frac{\partial}{\partial s}V_i^{(\boldsymbol{e}_m)}(s,t)
\end{aligned}
\tag{6.5.5}
$$

Then insert the expression for $\frac{\partial}{\partial s}V_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t)$ obtained in Example 6.5.6 as well as the Thiele differential equations for the state-wise prospective reserves, given by

$$
\frac{\partial}{\partial s} V_i^{(\boldsymbol{e}_\ell)}(s,t) = r(s)V_i^{(\boldsymbol{e}_\ell)}(s,t) - b_i^\ell(s) - \sum_{j:j\neq i}\mu_{ij}(s)R_{ij}^\ell(s),
$$

and analogously for $V_i^{(\boldsymbol{e}_m)}$. By gathering the relevant terms, we obtain the desired differential equation. The boundary condition follows directly from (6.5.4). ☐

Remark 6.5.8. Note that the differential equation of Corollary 6.5.7 corresponds to a Thiele differential equation with interest rate $2r$, no sojourn payments and transition payments $R_{ij}^\ell R_{ij}^m$ from state i to state j. Consequently, the conditional covariance $m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}$ has the representation

$$
\begin{aligned}
m_i^{(\boldsymbol{e}_\ell+\boldsymbol{e}_m)}(s,t) &= \mathbb{E}\bigg[\int_s^t e^{-\int_s^u 2r(v)\,\mathrm{d}v}\,\mathrm{d}\widetilde{B}(u)\,\bigg|\, Z(s)=i\bigg] \\
\mathrm{d}\widetilde{B}(u) &= \sum_{i\in\mathcal{J}} \mathbb{1}_{(Z(u-)=i)} \sum_{j:j\neq i} R_{ij}^\ell(u)R_{ij}^m(u)\,\mathrm{d}N_{ij}(u).
\end{aligned}
$$

In the case of the conditional variance, i.e. $\ell = m$, these type of representations are known as Hattendorff's theorem, which has been formulated in a multi-state Markovian framework by Ramlau-Hansen (1988) using martingale techniques and further generalized in Norberg (1992). The result here may thus reveal that the probabilistic structures leading to Hattendorff's theorem is not only limited to the variance, but can be carried over to the covariance. △

From the conditional covariances (and variances when $\ell = m$), we are then able to put up the conditional covariance matrix of $\boldsymbol{U}(s,t)$ given $Z(s) = i$,

$$\boldsymbol{\Sigma}_i(s,t) = \left\{ m_i^{(\boldsymbol{e}_\ell + \boldsymbol{e}_m)} \right\}_{\ell,m=1,\dots,n} \tag{6.5.6}$$

as well as the corresponding correlation matrix

$$\boldsymbol{\rho}_i(s,t) = \left\{ \frac{m_i^{(\boldsymbol{e}_\ell + \boldsymbol{e}_m)}(s,t)}{\sqrt{m_i^{(\boldsymbol{e}_\ell + \boldsymbol{e}_\ell)}(s,t) m_i^{(\boldsymbol{e}_m + \boldsymbol{e}_m)}(s,t)}} \right\}_{\ell,m=1,\dots,n} \tag{6.5.7}$$

which makes us able to analyze pair-wise dependence structures between the present values. In particular, these may be used to approximate joint safety margins via multivariate CLT approximations as outlined in Section 6.2.

6.5.2 Product integral representation of conditional partial moments

In this subsection, we derive a product integral representation for the conditional partial moments. This allows for the partial moments to be treated both theoretically and numerically as objects of its own within the theory of product integrals, where all properties and results for these can be benefited from, see e.g. Johansen (1986) and Gill and Johansen (1990) for theoretical properties and Helton and Stuckwisch (1976) for numerical schemes. Specific to this problem, the product integral shall even demonstrate how one may compute all moments $\boldsymbol{V}^{(\boldsymbol{y})}$, $\boldsymbol{y} \in \widetilde{S}(\boldsymbol{k})$, at once.

From the differential equations for the conditional (partial) moments, we see a structure similar to those of a single payment stream, namely that one must use all lower order moments $\boldsymbol{y}^1, \dots, \boldsymbol{y}^{|\boldsymbol{k}|-1} \in S(\boldsymbol{k})$ (each computed using Theorem 6.5.1 or 6.5.3) to compute the $\boldsymbol{k}$'th moment. Since each of these need their corresponding lower order moments $S(\boldsymbol{y}^m)$, $m = 1, \dots, |\boldsymbol{k}| - 1$, we are able to compute the $\boldsymbol{k}$'th moment starting from the $\boldsymbol{0}$'th moment, which is $\boldsymbol{V}^{(\boldsymbol{0})}(s,t) = \boldsymbol{P}(s,t)$ (or $V_i^{(\boldsymbol{0})}(s,t) = 1$), and then iteratively for each $m = 1, \dots, |\boldsymbol{k}|$ compute the $\boldsymbol{y}^m$'th moment using the lower order moments $\boldsymbol{y}^1, \dots, \boldsymbol{y}^{m-1}$. In total, one must then solve $(|\boldsymbol{k}| + 1)$ $J \times J$-dimensional systems of differential equations for $\boldsymbol{V}^{(\boldsymbol{k})}$, and $(|\boldsymbol{k}| + 1)$ J-dimensional systems for $V_i^{(\boldsymbol{k})}$.

However, this approach relies on how we sort the vectors $\boldsymbol{y}^1, \dots, \boldsymbol{y}^{|\boldsymbol{k}|}$, or, equivalently, the order at which we solve the differential equations, since we must ensure that the sets $S(\boldsymbol{y}^m)$ are increasing, i.e. that $S(\boldsymbol{y}^{m-1}) \subseteq S(\boldsymbol{y}^m)$ for all $m = 1, \dots, |\boldsymbol{k}|$, such that we actually are able to draw upon all lower order moments when calculating a given moment $\boldsymbol{y}^m$. Note that this is a trivial matter in the case of a single payment process, since for $n = 1$ we have $S(k) = \{1, \dots, k\}$.

It turns out that the key to obtain the result is to order the vectors $\boldsymbol{y}^1, \dots, \boldsymbol{y}^{|\boldsymbol{k}|}$ in a way that fortunately is standard in most software packages.

Assumption 6.5.9. *We assume that the lower order moments in the set* $S(\boldsymbol{k}) = \{\boldsymbol{y}^1, \dots, \boldsymbol{y}^{|\boldsymbol{k}|}\}$ *are lexicographical ordered, that is, they satisfy that for all* $m' > m$, $m, m' \in \{1, \dots, |\boldsymbol{k}|\}$, *there exist* $u \in \{1, \dots, n\}$ *(dependent on* m *and* m'*), such that*

$$\begin{aligned} &y_\ell^m = y_\ell^{m'} \text{ for all } \ell < u, \text{ and} \\ &y_u^m < y_u^{m'}. \end{aligned}$$

In other words, $\boldsymbol{y}^{m'}$ is strictly larger than $\boldsymbol{y}^m$ in the first entry where they differ (it may be smaller in the remaining entries). Details regarding this type of ordering and its properties are presented in Appendix 6.A.

Now, define the matrix $\boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}(x)$, $x \geq 0$, as

$$\begin{pmatrix} \boldsymbol{M}(x) - \bar{\boldsymbol{k}} r(x)\boldsymbol{I} & \boldsymbol{f}_1^{(\boldsymbol{k})}(x) & \boldsymbol{f}_2^{(\boldsymbol{k})}(x) & \cdots & \boldsymbol{f}_{|\boldsymbol{k}|}^{(\boldsymbol{k})}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x) - \bar{\boldsymbol{y}}^{|\boldsymbol{k}|-1} r(x)\boldsymbol{I} & \boldsymbol{f}_1^{(\boldsymbol{y}^{|\boldsymbol{k}|-1})}(x) & \cdots & \boldsymbol{f}_{|\boldsymbol{k}|-1}^{(\boldsymbol{y}^{|\boldsymbol{k}|-1})}(x) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) - \bar{\boldsymbol{y}}^1 r(x)\boldsymbol{I} & \boldsymbol{f}_1^{(\boldsymbol{y}^1)}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix}$$

with $\boldsymbol{f}_u^{(\boldsymbol{y}^m)}(x)$, for $m = 1, \dots, |\boldsymbol{k}|$ and $u = 1, \dots, m$, given by, setting $\boldsymbol{y}^0 := \boldsymbol{0}$,

$$\boldsymbol{f}_u^{(\boldsymbol{y}^m)}(x) = \begin{cases} y_1^m \boldsymbol{R}_1(x) & \text{if } \boldsymbol{y}^m - \boldsymbol{y}^{m-u} = \boldsymbol{e}_1, \\ y_2^m \boldsymbol{R}_2(x) & \text{if } \boldsymbol{y}^m - \boldsymbol{y}^{m-u} = \boldsymbol{e}_2, \\ \vdots & \vdots \\ y_n^m \boldsymbol{R}_n(x) & \text{if } \boldsymbol{y}^m - \boldsymbol{y}^{m-u} = \boldsymbol{e}_n, \\ \prod_{\ell=1}^n \binom{y_\ell^m}{\xi_\ell} \boldsymbol{C}^{(\boldsymbol{\xi})}(x), & \text{if } \boldsymbol{y}^m - \boldsymbol{y}^{m-u} = \boldsymbol{\xi} \in S(\boldsymbol{y}^m) \setminus E_n, \\ \boldsymbol{0} & \text{Otherwise.} \end{cases} \tag{6.5.8}$$

Here, one should think of the matrices $\left\{ \boldsymbol{f}_u^{(\boldsymbol{y}^m)}(x) \right\}_{u=1}^m$ as those needed to compute the $\boldsymbol{y}^m$'th moment, cf. Theorem 6.5.1, and so each block row of $\boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}(x)$ corresponds to a calculation of a lower order moment.

Example 6.5.10 (Conditional covariance)**.** For the computation of conditional covariances, we have for $n = 2$ and $\boldsymbol{k} = (1, 1)$ that the matrix $\boldsymbol{F}_{\boldsymbol{U}}^{(1,1)}(x)$ reads as

$$\boldsymbol{F}_{\boldsymbol{U}}^{(1,1)}(x) = \begin{pmatrix} \boldsymbol{M}(x) - 2r(x)\boldsymbol{I} & \boldsymbol{R}_2(x) & \boldsymbol{R}_1(x) & \boldsymbol{C}^{(1,1)}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{R}_1(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{R}_2(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix},$$

arising from the fact that the lexicographical ordering of $S((1, 1))$ is given as $(0, 1), (1, 0)$, and $(1, 1)$. ◦

Now, let

$$\boldsymbol{G}^{(\boldsymbol{k})}(s, t) = \prod_s^t \left(\boldsymbol{I} + \boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}(x)\mathrm{d}x\right) \tag{6.5.9}$$

denote the product integral on the interval $(s, t]$ of the matrix function $\boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}$. Since the matrix $\boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}(x)$ is a $\big(|\boldsymbol{k}|+1\big) \times \big(|\boldsymbol{k}|+1\big)$ block-partitioned matrix with blocks of sizes $J \times J$, we have that $\boldsymbol{G}^{(\boldsymbol{k})}(s, t)$ is of similar size, and we denote with $\boldsymbol{G}_{ij}^{(\boldsymbol{k})}(s, t)$ the ij'th block of $\boldsymbol{G}^{(\boldsymbol{k})}(s, t)$ for $i, j \in \{1, \dots |\boldsymbol{k}| + 1\}$. The following result then demonstrates how all moments of up to order $\boldsymbol{k}$ are obtained through this single calculation of a product integral.

Theorem 6.5.11. *For each $i \in \{0, 1, \dots, |\boldsymbol{k}|\}$ we have, for $m = 0, 1, \dots, i$,*

$$\boldsymbol{G}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1-m}^{(\boldsymbol{k})}(s, t) = \mathbb{1}_{(\boldsymbol{y}^i \geq \boldsymbol{y}^m)} \prod_{\ell=1}^n \binom{y_\ell^i}{y_\ell^m} e^{-\bar{\boldsymbol{y}}^m \int_s^t r(v)\,\mathrm{d}v} \boldsymbol{V}^{(\boldsymbol{y}^i - \boldsymbol{y}^m)}(s, t), \tag{6.5.10}$$

which in particular gives the conditional $\boldsymbol{y}^i$'th moment for $m = 0$:

$$\boldsymbol{G}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1}^{(\boldsymbol{k})}(s, t) = \boldsymbol{V}^{(\boldsymbol{y}^i)}(s, t). \tag{6.5.11}$$

Proof. We shall mimic the proof of Bladt, Asmussen, and Steffensen (2020, Theorem 3-4) and modify suitably to the present setup. By multiplying (6.5.1) with $\prod_t^s \big(\boldsymbol{I} + \big[\boldsymbol{M}(u) - \bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)$ on both sides we get

$$\begin{aligned}
&\frac{\partial}{\partial s}\left(\prod_t^s \big(\boldsymbol{I} + \big[\boldsymbol{M}(u) - \bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)\boldsymbol{V}^{(\boldsymbol{k})}(s, t)\right) \\
&= -\sum_{\ell=1}^n k_\ell \prod_t^s \big(\boldsymbol{I} + \big[\boldsymbol{M}(u) - \bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)\boldsymbol{R}_\ell(s)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{e}_\ell)}(s, t) \\
&\quad - \sum_{\substack{\boldsymbol{y}\in S(\boldsymbol{k}) \\ \boldsymbol{y}\notin E_n}} \prod_{\ell=1}^n \binom{k_\ell}{y_\ell} \prod_t^s \big(\boldsymbol{I} + \big[\boldsymbol{M}(u) - \bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)\boldsymbol{C}^{(\boldsymbol{y})}(s)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(s, t).
\end{aligned}$$

Integrating the equation yields

$$
\begin{aligned}
&\boldsymbol{V}^{(\boldsymbol{k})}(s,t)\\
&=\sum_{\ell=1}^{n} k_\ell \int_s^t \prod_s^x \big(\boldsymbol{I}+\big[\boldsymbol{M}(u)-\bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)\boldsymbol{R}_\ell(x)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{e}_\ell)}(x,t)\,\mathrm{d}x\\
&+\sum_{\substack{\boldsymbol{y}\in S(\boldsymbol{k})\\ \boldsymbol{y}\notin E_n}} \prod_{\ell=1}^{n}\binom{k_\ell}{y_\ell}\int_s^t \prod_s^x \big(\boldsymbol{I}+\big[\boldsymbol{M}(u)-\bar{\boldsymbol{k}}r(u)\boldsymbol{I}\big]\mathrm{d}u\big)\boldsymbol{C}^{(\boldsymbol{y})}(x)\boldsymbol{V}^{(\boldsymbol{k}-\boldsymbol{y})}(x,t)\,\mathrm{d}x.
\end{aligned}
$$

We now provide an induction argument to verify the identity (6.5.10) claimed in the theorem using this integral equation. For $i=0$ and $m=0$ the identity is trivially true. So assume that the identity is true for some $i-1$, $i\in\{1,\dots,|\boldsymbol{k}|+1\}$, and corresponding $m=0,1\dots,i-1$. Then Bladt, Asmussen, and Steffensen (2020, Lemma 1) gives us that

$$
\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1-i}(s,t)=\prod_s^t\big(\boldsymbol{I}+\big[\boldsymbol{M}(x)-\bar{\boldsymbol{y}}^i r(x)\boldsymbol{I}\big]\mathrm{d}x\big)=e^{-\bar{\boldsymbol{y}}^i\int_s^t r(v)\,\mathrm{d}v}\boldsymbol{P}(s,t),
$$

which is (6.5.9) for $m=i$. From the Lemma, it also follows that for $m=0,1,\dots,i-1$,

$$
\begin{aligned}
&\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1-m}(s,t)\\
&=\sum_{j=1}^{i-m}\int_s^t e^{-\bar{\boldsymbol{y}}^i\int_s^x r(v)\,\mathrm{d}v}\boldsymbol{P}(s,x)\boldsymbol{f}_j^{(\boldsymbol{y}^i)}(x)\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-(i-j),|\boldsymbol{k}|+1-m}(x,t)\mathrm{d}x
\end{aligned}
$$

From the induction hypothesis we have that

$$
\begin{aligned}
&\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-(i-j),|\boldsymbol{k}|+1-m}(x,t)\\
&=\mathbb{1}_{(\boldsymbol{y}^{i-j}\geq\boldsymbol{y}^m)}\prod_{\ell=1}^{n}\binom{y_\ell^{i-j}}{y_\ell^m}e^{-\bar{\boldsymbol{y}}^m\int_x^t r(v)\,\mathrm{d}v}\boldsymbol{V}^{(\boldsymbol{y}^{i-j}-\boldsymbol{y}^m)}(x,t),
\end{aligned}
$$

and furthermore, $\boldsymbol{f}_j^{(\boldsymbol{y}^i)}(x)$ may be written as

$$
\boldsymbol{f}_j^{(\boldsymbol{y}^i)}(x)=\sum_{\ell=1}^{n}\mathbb{1}_{(\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{e}_\ell)}y_\ell^i\boldsymbol{R}_\ell(x)+\sum_{\substack{\boldsymbol{\xi}\in S(\boldsymbol{y}^i)\\ \boldsymbol{\xi}\notin E_n}}\mathbb{1}_{(\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{\xi})}\prod_{\ell=1}^{n}\binom{y_\ell^i}{\xi_\ell}\boldsymbol{C}^{(\boldsymbol{\xi})}(x).
$$

Note that when $\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{\xi}$ for some $\boldsymbol{\xi}\in S(\boldsymbol{y}^i)$ (including the unit vectors) and some $j=1,\dots,i-m$, it holds

$$
\binom{y_\ell^{i-j}}{y_\ell^m}\binom{y_\ell^i}{\xi_\ell}=\binom{y_\ell^i}{y_\ell^m}\binom{y_\ell^i-y_\ell^m}{\xi_\ell}
$$

for all $\ell=1,\dots,n$. Thus, we now have

$$
\begin{aligned}
&\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1-m}(s,t)\\
&= \prod_{\ell=1}^{n}\binom{y^i_\ell}{y^m_\ell}\int_s^t e^{-\bar{\boldsymbol{y}}^i\int_s^x r(v)\,\mathrm{d}v}\boldsymbol{P}(s,x)e^{-\bar{\boldsymbol{y}}^m\int_x^t r(v)\,\mathrm{d}v}\sum_{j=1}^{i-m}\mathbb{1}_{(\boldsymbol{y}^{i-j}\geq\boldsymbol{y}^m)}\times\\
&\Bigg\{\sum_{\ell=1}^{n}\mathbb{1}_{(\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{e}_\ell)}(y^i_\ell-y^m_\ell)\boldsymbol{R}_\ell(x)\boldsymbol{V}^{(\boldsymbol{y}^i-\boldsymbol{y}^m-\boldsymbol{e}_\ell)}(x,t)\\
&\quad+\sum_{\substack{\boldsymbol{\xi}\in S(\boldsymbol{y}^i)\\ \boldsymbol{\xi}\notin E_n}}\mathbb{1}_{(\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{\xi})}\prod_{\ell=1}^{n}\binom{y^i_\ell-y^m_\ell}{\xi_\ell}\boldsymbol{C}^{(\boldsymbol{\xi})}(x)\boldsymbol{V}^{(\boldsymbol{y}^i-\boldsymbol{y}^m-\boldsymbol{\xi})}(x,t)\Bigg\}\mathrm{d}x
\end{aligned}
$$

Then observe that the discount factors can be factored as, for $x\in[s,t]$,

$$
e^{-\bar{\boldsymbol{y}}^i\int_s^x r(v)\,\mathrm{d}v}e^{-\bar{\boldsymbol{y}}^m\int_x^t r(v)\,\mathrm{d}v}=e^{-\bar{\boldsymbol{y}}^m\int_s^t r(v)\,\mathrm{d}v}e^{-(\bar{\boldsymbol{y}}^i-\bar{\boldsymbol{y}}^m)\int_s^x r(v)\,\mathrm{d}v}.
$$

Furthermore, when $\boldsymbol{y}^{i-j}\geq\boldsymbol{y}^m$ and $\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{\xi}$ for some $\boldsymbol{\xi}\in S(\boldsymbol{y}^i)$ and $j=1,\ldots,i-m$, we have that $\boldsymbol{\xi}\leq\boldsymbol{y}^i-\boldsymbol{y}^m$, and so the last sum can be carried out over $S(\boldsymbol{y}^i-\boldsymbol{y}^m)\setminus E_n$. Thus, we now have

$$
\begin{aligned}
&\boldsymbol{G}^{(\boldsymbol{k})}_{|\boldsymbol{k}|+1-i,|\boldsymbol{k}|+1-m}(s,t)\\
&\quad= \mathbb{1}_{(\boldsymbol{y}^i\geq\boldsymbol{y}^m)}\prod_{\ell=1}^{n}\binom{y^i_\ell}{y^m_\ell}e^{-\bar{\boldsymbol{y}}^m\int_s^t r(v)\,\mathrm{d}v}\times\\
&\qquad\qquad\int_s^t e^{-(\bar{\boldsymbol{y}}^i-\bar{\boldsymbol{y}}^m)\int_s^x r(v)\,\mathrm{d}v}\boldsymbol{P}(s,x)\Bigg\{\sum_{\ell=1}^{n}(y^i_\ell-y^m_\ell)\boldsymbol{R}_\ell(x)\boldsymbol{V}^{(\boldsymbol{y}^i-\boldsymbol{y}^m-\boldsymbol{e}_\ell)}(x,t)\\
&\qquad+\sum_{\substack{\boldsymbol{\xi}\in S(\boldsymbol{y}^i-\boldsymbol{y}^m)\\ \boldsymbol{\xi}\notin E_n}}\prod_{\ell=1}^{n}\binom{y^i_\ell-y^m_\ell}{\xi_\ell}\boldsymbol{C}^{(\boldsymbol{\xi})}(x)\boldsymbol{V}^{(\boldsymbol{y}^i-\boldsymbol{y}^m-\boldsymbol{\xi})}(x,t)\Bigg\}\mathrm{d}x\\
&= \mathbb{1}_{(\boldsymbol{y}^i\geq\boldsymbol{y}^m)}\prod_{\ell=1}^{n}\binom{y^i_\ell}{y^m_\ell}e^{-\bar{\boldsymbol{y}}^m\int_s^t r(v)\,\mathrm{d}v}\boldsymbol{V}^{(\boldsymbol{y}^i-\boldsymbol{y}^m)}(s,t),
\end{aligned}
$$

as claimed. In the first equality, we have used that for all $\boldsymbol{\xi}\in S(\boldsymbol{y}^i-\boldsymbol{y}^m)$,

$$
\sum_{j=1}^{i-m}\mathbb{1}_{(\boldsymbol{y}^i-\boldsymbol{y}^{i-j}=\boldsymbol{\xi})}=\mathbb{1}_{(\boldsymbol{y}^i\geq\boldsymbol{y}^m)},
$$

which follows from Lemma 6.A.1. $\square$

The theorem gives us that the right-block column of $\boldsymbol{G}^{(\boldsymbol{k})}(s,t)$ contains all moments of order up to $\boldsymbol{k}$, including the $\boldsymbol{0}$'th moment, as follows:

$$\prod_s^t \left(\boldsymbol{I} + \boldsymbol{F}_{\boldsymbol{U}}^{(\boldsymbol{k})}(x)\,\mathrm{d}x\right) = \begin{pmatrix} * & * & * & * & \dots & * & \boldsymbol{V}^{(\boldsymbol{k})}(s,t) \\ * & * & * & * & \dots & * & \boldsymbol{V}^{(\boldsymbol{y}^{|\boldsymbol{k}|-1})}(s,t) \\ * & * & * & * & \dots & * & \boldsymbol{V}^{(\boldsymbol{y}^{|\boldsymbol{k}|-2})}(s,t) \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ * & * & * & * & \dots & * & \boldsymbol{V}^{(\boldsymbol{y}^{1})}(s,t) \\ * & * & * & * & \dots & * & \boldsymbol{P}(s,t) \end{pmatrix},$$

and so all lower order up to order $\boldsymbol{k}$ are obtained through this product integral calculation. From this, the conditional moments $V_i^{(\boldsymbol{k})}$ and corresponding central moments $m_i^{(\boldsymbol{k})}$ are obtained via (6.3.6) and (6.5.3), respectively.

6.6 Numerical example

In this section, we illustrate the methods presented in the previous sections in a numerical example of the motivating example considered in Section 2.

We consider a 40-year old male today at time 0 with retirement age 65, such that $T = 25$. We suppose that the valuation basis is taken to be the technical basis in the numerical example of Buchardt and Møller (2015), which is given by the following:

$$\begin{aligned} r(s) &= 0.01, \\ \mu_{01}(s) &= \left(0.0004 + 10^{4.54+0.06(s+40)-10}\right)\mathbb{1}_{(s\le 25)}, \\ \mu_{10}(s) &= \left(2.0058 \cdot e^{-0.117(s+40)}\right)\mathbb{1}_{(s\le 25)}, \\ \mu_{02}(s) &= 0.0005 + 10^{5.88+0.038(s+40)-10}, \\ \mu_{12}(s) &= \mu_{02}(s)\left(1 + \mathbb{1}_{(s\le 25)}\right). \end{aligned}$$

Some of the transition intensities are inspired by the Danish G82M technical basis. We examine the pair-wise covariance and correlation structures of the product combinations in the active state, i.e. the entries of the conditional covariance and correlation matrix given in (6.5.6)–(6.5.7) with $i = 0$; this is illustrated in Figure 6.2–6.3. The calculations are based on a numerical solution of the differential equation of Corollary 6.5.7 for the conditional covariance.

We see that the biggest dependence is between the life annuity and death benefit, which is was expected, but we also see a similar (but opposite) dependence structure between the disability annuity and death benefit respectively life annuity. The death benefit seem to be slightly more correlated with the disability annuity than the life annuity is, which is the most non-trivial observation encountered here. The calculated covariances can further be used to carry out the CLT approximation of the joint distribution outlined in Section 6.2, which we refrain from doing here.

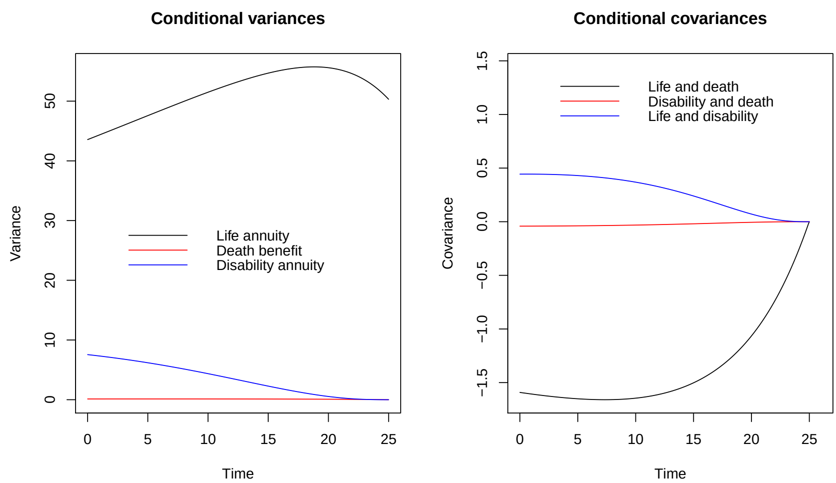

Figure 6.2: *Pair-wise conditional covariances* $[0,25] \in s \mapsto \Sigma_0(s,70)_{\ell m}$, $\ell, m \in \{1,2,3\}$, $\ell \leq m$, *between the three products until retirement conditional on the insured being active.*

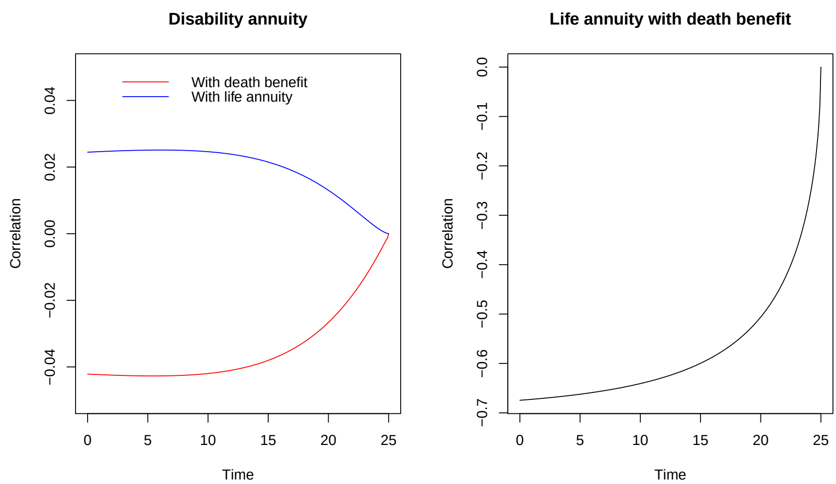

Figure 6.3: *Pair-wise conditional correlations* $[0,25] \in s \mapsto \rho_0(s,70)_{\ell m}$, $\ell, m \in \{1,2,3\}$, $\ell < m$, *between the three products until retirement conditional on the insured being active.*

Acknowledgments

I should like to thank my PhD supervisors Mogens Bladt and Mogens Steffensen for valuable suggestions and fruitful discussions during this work.

6.A Lexicographical ordering

Let $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{N}_0^n$ be a n-dimensional vector of natural numbers including zero. We think of this vector as representing some multivariate higher order moment we wish to calculate for the multivariate present value $\boldsymbol{U}$. Specific to this, we need to consider all combinations of lower order moments represented by the elements of $S(\boldsymbol{k})$; recall (6.3.1) for its definition. Write these elements as

$$S(\boldsymbol{k}) = \left\{\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{|\boldsymbol{k}|}\right\}.$$

The vectors $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{|\boldsymbol{k}|}$ are then said to be *lexicographical ordered* if for all $m' > m$, there exist u (dependent on m and m') such that

$$\begin{aligned} & y_\ell^m = y_\ell^{m'} \text{ for all } \ell < u, \text{ and} \\ & y_u^m < y_u^{m'}. \end{aligned}$$

In other words $y^{m'}$ must be strictly larger than y^m in the first entry where they differ (it may be smaller in the remaining entries), and one commonly writes $\boldsymbol{y}^m <_{\text{lex}} \boldsymbol{y}^{m'}$. An illustration of the ordering is presented in Table 6.1.

$\boldsymbol{y}^1 = (0, \ldots, 0, 1)'$	$\boldsymbol{y}^{k_n+1} = (0, \ldots, 1, 0)$		$\boldsymbol{y}^{|\boldsymbol{k}|-k_n} = (k_1, \ldots, k_{n-1}, 0)'$
$\boldsymbol{y}^2 = (0, \ldots, 0, 2)'$	$\boldsymbol{y}^{k_n+2} = (0, \ldots, 1, 1)$		$\boldsymbol{y}^{|\boldsymbol{k}|-k_n+1} = (k_1, \ldots, k_{n-1}, 1)'$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
$\boldsymbol{y}^{k_n} = (0, \ldots, 0, k_n)'$	$\boldsymbol{y}^{2(k_n+1)-1} = (0, \ldots, 1, k_n)'$		$\boldsymbol{y}^{|\boldsymbol{k}|} = (k_1, \ldots, k_n)'$

Table 6.1: *Illustration of lexicographical ordering of vectors in $S(\boldsymbol{k})$.*

The following result on this type of ordering provides the foundation to successively compute multivariate higher order moments based on already computed lower order moments.

Theorem 6.A.1. *Assume that $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{|\boldsymbol{k}|}$ are lexicographical ordered. Then for all $i = 1, \ldots, |\boldsymbol{k}|$ and $m = 0, 1, \ldots, i-1$ such that $\boldsymbol{y}^i \geq \boldsymbol{y}^m$, we have*

$$S(\boldsymbol{y}^i - \boldsymbol{y}^m) \subseteq \bigcup_{j=m}^{i-1} \left\{\boldsymbol{y}^i - \boldsymbol{y}^j\right\}.$$

Sketch of proof. We show the result for the two-dimensional case $n = 2$, and the generalization to higher dimension then follows using same (but notationally cumbersome) techniques. Let $i = 1, \ldots, |\boldsymbol{k}|$ be given. From the illustration of the lexicographic ordering of vectors in $S(\boldsymbol{k})$ shown in Table 6.1, we may realize that the vectors take the form

$$\boldsymbol{y}^i = (a, i - a(k_2+1))', \quad \text{if } i \in [a(k_2+1), (a+1)(k_2+1) - 1]$$

where $a = 1, \ldots, k_1$. This can be verified by exploiting the structure of which the ordering is carried out. Now let $m = 0, 1 \ldots, i-1$ be given such that $\boldsymbol{y}^i \geq \boldsymbol{y}^m$. Then similarly there exist $b = 0, 1, \ldots, k_1$ such that $m \in [b(k_2+1), (b+1)(k_2+1)-1]$, which then gives

$$\boldsymbol{y}^i - \boldsymbol{y}^m = (a-b, i-m-(a-b)(k_2+1))', \tag{6.A.1}$$

By similar reasonings, we have for $j = m, \ldots, i-1$ there exist $c = 1, \ldots, k_1$ such that $j \in [c(k_2+1), (c+1)(k_2+1)-1]$, giving

$$\boldsymbol{y}^i - \boldsymbol{y}^j = (a-c, i-j-(a-c)(k_2+1))'.$$

Due to the indices having the order $m \leq j \leq i$, we have that $b \leq c \leq a$. Consequently, the set $\bigcup_{j=m}^{i-1} \left\{\boldsymbol{y}^i - \boldsymbol{y}^j\right\}$ consists of vectors on the form

$$(a-c, i-j-(a-c)(k_2+1))'$$

where j varies on $m, \ldots, i-1$ and c varies on $b, b+1, \ldots, a$.

Now take $\boldsymbol{\xi} \in S(\boldsymbol{y}^i - \boldsymbol{y}^m)$. Since $\boldsymbol{\xi} \leq \boldsymbol{y}^i - \boldsymbol{y}^m$, it follows from (6.A.1) that there exist $d \leq a-b$ and $r \leq i-m-(a-b)(k_2+1)$, $d, r \in \mathbb{N}_0$, such that

$$\begin{aligned}
\boldsymbol{\xi} &= (a-(b+d), i-(m+r)-(a-b)(k_2+1))' \\
&= (a-(b+d), i-(m+r+d(k_2+1))-(a-(b+d))(k_2+1))',
\end{aligned}$$

where we have added and subtracted the term $d(k_2+1)$ in the second coordinate to obtain the last equality. We now see that $\boldsymbol{\xi}$ is on the form

$$(a-c, i-j-(a-c)(k_2+1))'$$

with $c = b+d$ and $j = m+r+d(k_2+1)$, and so we have $\boldsymbol{\xi} \in \bigcup_{j=m}^{i-1} \left\{\boldsymbol{y}^i - \boldsymbol{y}^j\right\}$, as claimed. This concludes the sketch of the proof. $\square$

Chapter 7

Computation of bonus in multi-state life insurance

This chapter is based on the paper *Ahmad, Buchardt, and Furrer (2022).*

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 classic 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. Methods and results are illustrated via a numerical example.

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

7.1 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 classic analytical methods and numerical methods for differential equations in regards to the outstanding insurance risk.

In the preexisting literature on valuation of both with-profit and equity/unit-linked life insurance, the focus is prevalently on financial risk, so that it is not uncommon to disregard biometric and behavioral risks more or less completely, see e.g. Bauer et al. (2006) and Zaglauer and Bauer (2008) and, concerning with-profit life insurance, cf. the discussion in Jensen and Schomacker (2015, Section 1). Exceptions include for instance Bacinello (2001) and Møller and Steffensen (2007) and, more recently, Bacinello, Millossovich, and Chen (2018), where both financial risk and mortality/longevity risk are considered. Our emphasis is in the spirit of Jensen and Schomacker (2015) and concerns the specific challenges that universally arise from including event risk via multi-state modeling. This places our research as part of the literature on multi-state modeling for with-profit life insurance. For a non-technical introduction to multi-state modeling in life insurance, see Koller (2012).

For Danish with-profit products, 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, cf. Møller and Steffensen (2007, Chapter 2). This is achieved by considering the available assets together with the market value of guaranteed payments and imposing the equivalence principle on the market basis in combination with certain ad hoc adjustments. In case parts of the assets correspond to expected future profits, also denoted as contractual service margins in the IFRS 17 regulatory framework, cf. Board (2017), the equivalence principle is invalidated, which points to more sophisticated computational methods. The provision of these kinds of methods constitutes the main contribution of this paper.

We distinguish between dividends and bonuses: The development of the portfolio and the financial market typically gives rise to a surplus, which is distributed among the policyholders' policies via dividend yields according to the chosen dividend strategy. After this allocation, the accumulated dividends on each policy are paid out according to a specific bonus scheme, and we refer to these extra payments as the bonus payments. The study of systematic surplus, dividends 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 (1995a). 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 (2021), and Falden and Nyegaard (2021). 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 portfolio of insured and although decisions by the insurer (so-called future management actions), including possible determination of dividend yields, often depend mainly on the performance of said portfolio, one ought to take into account that bonus payments in principle are allocated to the individual insured. This is the starting point in Bruhn and Lollike (2021), 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 (2021), the results of Bruhn and Lollike (2021) are extended to allow for policyholder behavior, namely the options of surrender and free policy conversion. The surrender option allows the policyholder to cancel all future payments and instead receive a single payment corresponding in some sense to the value of the contract, while the free policy option allows the policyholder to cancel future premiums at the cost of reducing future benefits. In Bruhn and Lollike (2021) and Falden and Nyegaard (2021), the dependence of the dividend strategy on the financial state of the insurance business, encapsulated in what we below shall term the *shape* of the insurance business, and the practical and computational challenges arising from this are not highlighted nor studied. This paper derives its main novelty value from addressing these challenges within a multi-state framework while also allowing for financial risk.

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 incidental modeling of the policyholder options surrender and free policy conversion following Henriksen et al. (2014), Buchardt and Møller (2015), and Buchardt, Møller, and Schmidt (2015). In regards to dividends and bonus, we adopt a somewhat universal dividend strategy; examples throughout the text, including the numerical example, provide links to the preexisting literature and actuarial practice by focusing on dividends arising via a second order interest rate, which in particular gives rise to an interest rate guarantee. The allocated dividends are later paid out according to a bonus scheme, and we here focus solely on the bonus scheme known as *additional benefits*, where dividends are simply used to buy extra benefits; this bonus scheme constitutes the focal point of Møller and Steffensen (2007, Chapter 6) and is also quite common in practice.

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 first and foremost 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 and thus are relevant to the insurer in determining the dividend strategy and investment strategy. In particular, the shape of the insurance business includes portfolio-wide means of the technical reserve of guaranteed payments and of the expected accumulated guaranteed cash flows. 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. By not needing to simulate insurance risk, our procedure has a significant advantage compared to full-blown Monte Carlo methods. 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. This is further examined in a numerical example.

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 7.2, we present the setup. The general results and general numerical procedure are given in Section 7.3, while the state-independent case is the subject of Section 7.4. Section 7.5 is devoted to a numerical example. Finally, Section 7.6 concludes with a comparison to recent advances in the literature and a discussion of possible extensions.

7.2 Setup

In the following, we describe the mathematical framework. Subsections 7.2.1–7.2.3 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 7.2.4, 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 7.2.5. In the remainder of the paper, the problem is studied for a specific class of dividend processes specified in Subsection 7.2.6.

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.

7.2.1 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

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

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

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

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

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

with $\mathcal{J}^p := \{0, \ldots, J\}$ and $\mathcal{J}^f := \{J+1, \ldots, 2J+1\}$ for some $J \in \mathbb{N}$. Here $\mathcal{J}^p$ contains the premium paying states, while $\mathcal{J}^f$ 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 $\mathcal{J}^f$ 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 7.1.

7.2.2 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 $\mathcal{J}^p$ 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 the time of free policy conversion. In rigorous terms, we have

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

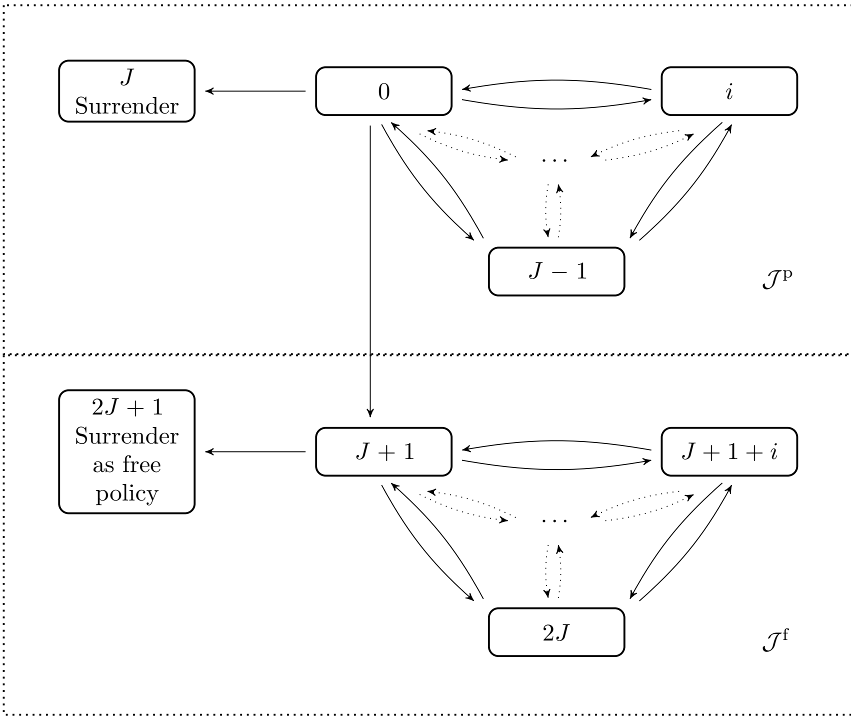

Figure 7.1: *General finite state space extended with a surrender state $\{J\}$ and free policy states $\mathcal{J}^f$. The states $\mathcal{J}^p \setminus \{J\}$ contain the biometric states of the insured, e.g. active, disabled, and dead. The states $\mathcal{J}^{\mathrm{f}}$ are a copy of $\mathcal{J}^p$, 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.*

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

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

We have $\tau = 0$ if and only if $z_0 \in \mathcal{J}^f$; 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 superscripts $\pm$ to denote the benefit and premium part, respectively. Then we have

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

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 7.2.4, are zero for $t > n$.

7.2.3 Valuation of the 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 7.1, are retained under $\mathrm{P}^\star$. Policyholder behavior is typically not explicitly included on the technical basis; in particular, there is no change in technical reserve upon free policy conversion. Following along the lines of Buchardt and Møller (2015) and Buchardt, Møller, and Schmidt (2015), this is consistent with the assumption that the transition rates under the technical basis, the surrender payments, and the free policy factor take the form

$$\begin{aligned} &\mu^\star_{jk} = \mu^\star_{j'k'}, && j,k \in \mathcal{J}^f, k \neq j, \\ &b_{0J} = \widetilde{V}^\star_0, && \\ &(0,\infty) \ni t \mapsto \rho(t) = \frac{\widetilde{V}^\star_0(t)}{\widetilde{V}^{\star,+}_0(t)}, && \end{aligned} \tag{7.2.1}$$

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

$$\widetilde{V}^\star_j(t) = \mathbb{E}^\star\left[\int_t^n e^{-\int_t^s r^\star(v)\,\mathrm{d}v}\,\mathrm{d}B^\circ(s)\,\middle|\, Z(t) = j\right], \tag{7.2.2}$$

$$\widetilde{V}^{\star,+}_j(t) = \mathbb{E}^\star\left[\int_t^n e^{-\int_t^s r^\star(v)\,\mathrm{d}v}\,\mathrm{d}B^{\circ,+}(s)\,\middle|\, Z(t) = j\right], \tag{7.2.3}$$

with $\mathbb{E}^\star$ denoting integration w.r.t. $\mathrm{P}^\star$. The specification of the technical reserve is circular, since the payments depend on the technical reserve, which again depends on the payments.The setup described here is closely related to the generic situation described in Christiansen and Djehiche (2020, Section 4.3), where circularity is resolved by a careful construction of the involved processes and using backward stochastic differential techniques; for instance, our specification (7.2.1) ensures that equation (4.8) in Christiansen and Djehiche (2020) holds.

It is possible to show that the state-wise technical reserves of predetermined payments satisfy the following differential equations of Thiele type:

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{V}_j^\star(t) = r^\star(t)\widetilde{V}_j^\star(t) - b_j(t) - \sum_{\substack{k\in\mathcal{J}^p\setminus\{J\}\\ k\neq j}} \big(b_{jk}(t) + \widetilde{V}_k^\star(t) - \widetilde{V}_j^\star(t)\big)\mu_{jk}^\star(t), \quad \widetilde{V}_j^\star(n) = 0, \tag{7.2.4}$$

for $j \in \mathcal{J}^p \setminus \{J\}$. 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 $j \in \mathcal{J}^p \setminus \{J\}$ to $j \in \mathcal{J}$ via

$$V_j^{\star,\circ}(t) = \begin{cases} \widetilde{V}_j^\star(t) & \text{if } j \in \mathcal{J}^p \setminus \{J\}, \\ \rho(\tau)\widetilde{V}_{j'}^{\star,+}(t) & \text{if } j \in \mathcal{J}^f \setminus \{2J+1\}, \\ 0 & \text{if } j \in \{J, 2J+1\}. \end{cases} \tag{7.2.5}$$

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. The technical reserve of predetermined payments $V^{\star,\circ}$ is now defined according to $V^{\star,\circ}(t) = V_{Z(t)}^{\star,\circ}(t)$. In particular, $V^{\star,\circ}(t)$ is a version of the conditional expectation

$$\mathbb{E}^\star\bigg[\int_t^n e^{-\int_t^s r^\star(u)\,\mathrm{d}u}\,\mathrm{d}B^\circ(s)\,\bigg|\,\mathcal{F}(t)\bigg].$$

In practice, the technical reserves are exactly computed according to (7.2.4) and (7.2.5). Here we provide a probablistic setup and specification that is consistent with this approach.

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

$$V^\circ(t) = \mathbb{E}\bigg[\int_t^n e^{-\int_t^s r(u)\,\mathrm{d}u}\,\mathrm{d}B^\circ(s)\,\bigg|\,\mathcal{F}(t)\bigg] = \int_t^n e^{-\int_t^s f(t,u)\,\mathrm{d}u} A^\circ(t,\mathrm{d}s), \tag{7.2.6}$$

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

$$A^\circ(t,s) = \mathbb{E}\big[B^\circ(s) - B^\circ(t)\,|\,\mathcal{F}^Z(t)\big]. \tag{7.2.7}$$

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

$$A^\circ(t,\mathrm{d}s) = \rho(\tau)\sum_{j\in\mathcal{J}^f} p_{Z(t)j}(t,s)\bigg(b_{j'}(s)^+ + \sum_{\substack{k\in\mathcal{J}^f\\ k\neq j}} b_{j'k'}(s)^+\mu_{jk}(s)\bigg)\mathrm{d}s, \tag{7.2.8}$$

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

$$
\begin{aligned}
A^\circ(t, \mathrm{d}s) = &\sum_{j \in \mathcal{J}^p} p_{Z(t)j}(t,s)\bigg(b_j(s) + \sum_{\substack{k \in \mathcal{J}^p \\ k \neq j}} b_{jk}(s)\mu_{jk}(s)\bigg)\mathrm{d}s \\
&+ \sum_{j \in \mathcal{J}^f} p^\rho_{Z(t)j}(t,s)\bigg(b_{j'}(s)^+ + \sum_{\substack{k \in \mathcal{J}^f \\ k \neq j}} b_{j'k'}(s)^+\mu_{jk}(s)\bigg)\mathrm{d}s
\end{aligned}
\tag{7.2.9}
$$

where the so-called ρ-modified transition probabilities p^ρ_{jk}, $j \in \mathcal{J}^p$ and $k \in \mathcal{J}$, are defined by $p^\rho_{jk}(t,s) = \mathbb{E}[\mathbb{1}_{(Z(s)=k)}\rho(\tau)^{\mathbb{1}_{(\tau \leq s)}} \mid Z(t) = j]$ and satisfy for $k \in \mathcal{J}^f$ so-called ρ-modified versions of Kolmogorov's forward differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} p^\rho_{jk}(t,s) &= \sum_{\substack{\ell \in \mathcal{J}^f \\ \ell \neq k}} p^\rho_{j\ell}(t,s)\mu_{\ell k}(s) + \mathbb{1}_{(k=J+1)} p_{j0}(t,s)\mu_{0k}(s)\rho(s) - p^\rho_{jk}(t,s)\mu_{k\bullet}(s), \\
p^\rho_{jk}(t,t) &= 0,
\end{aligned}
\tag{7.2.10}
$$

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

7.2.4 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 7.2.6, we specify the dividend strategy further. Classic examples include dividends distributed via a second order interest rate, see Example 7.2.7 below.

The dividends are allocated to the individual life insurance contract but not yet paid out; payout occurs at a possibly later point in time according to some specific bonus scheme. In the following, we adopt the bonus scheme known as *additional benefits*, that is 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 $\mathcal{J}^p$, $b_j^\dagger$ and $b_{jk}^\dagger$, are suitably regular non-negative deterministic functions with $b_J^\dagger \equiv 0$, while

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

where for $j \in \mathcal{J}^p \setminus \{J\}$ we denote by $\widetilde{V}_j^{\star,\dagger}$ the state-wise technical unit reserves of $B^\dagger$ given as (7.2.2) with B° replaced by $B^\dagger$. Again, these state-wise technical reserves satisfy differential equations of Thiele type, namely (7.2.4) with added superscripts †.

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

$$V_j^{\star,\dagger}(t) = \begin{cases} \widetilde{V}_j^{\star,\dagger}(t) & \text{if } j \in \mathcal{J}^p \setminus \{J\}, \\ \widetilde{V}_{j'}^{\star,\dagger}(t) & \text{if } j \in \mathcal{J}^f \setminus \{2J+1\}, \\ 0 & \text{if } j \in \{J, 2J+1\}, \end{cases} \tag{7.2.11}$$

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

$$A^\dagger(t, \mathrm{d}s) = a^\dagger(t,s)\,\mathrm{d}s, \tag{7.2.12}$$

$$a^\dagger(t,s) = \sum_{j\in\mathcal{J}} p_{Z(t)j}(t,s)\Big(b_j^\dagger(s) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b_{jk}^\dagger(s)\mu_{jk}(s)\Big). \tag{7.2.13}$$

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

$$A_i^\dagger(t, \mathrm{d}s) = a_i^\dagger(t,s)\,\mathrm{d}s, \tag{7.2.14}$$

$$a_i^\dagger(t,s) = \sum_{j\in\mathcal{J}} p_{ij}(t,s)\Big(b_j^\dagger(s) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b_{jk}^\dagger(s)\mu_{jk}(s)\Big). \tag{7.2.15}$$

The market unit reserve of $B^\dagger$ denoted $V^\dagger$ is given analogously to (7.2.6) but with the superscript † replacing ∘. Note that contrary to the market reserve of predetermined payments, it may be taken to only depend on $Z(t)$ at time t, since this is the case for the corresponding expected accumulated unit bonus cash flow. We write $V_{Z(t)}^\dagger(t)$ in place of $V^\dagger(t)$ to highlight this fact.

Having specified the unit bonus payment stream $B^\dagger$ as well as its technical value and associated expected accumulated cash flows, we are now in a position to finalize the specification of the bonus scheme. Thus, let $Q(t)$ denote the number of additional benefits, that is the number of benefit streams $B^\dagger$, held at time t. Since δ is used as a premium to buy $B^\dagger$ on the technical basis, we have that (cf. equation (4.11) in Norberg (2001))

$$\mathrm{d}Q(t) = \frac{\mathrm{d}D(t)}{V^{\star,\dagger}_{Z(t)}(t)} = \frac{\delta(t)}{V^{\star,\dagger}_{Z(t)}(t)}\,\mathrm{d}t, \quad Q(0) = 0. \tag{7.2.16}$$

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

$$\mathrm{d}B(t) = \mathrm{d}B^\circ(t) + Q(t)\,\mathrm{d}B^\dagger(t), \quad B(0) = 0. \tag{7.2.17}$$

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

$$B^g(t, \mathrm{d}s) = \mathrm{d}B^\circ(s) + Q(t)\,\mathrm{d}B^\dagger(s), \quad B^g(t,t) = B(t), \tag{7.2.18}$$

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)*.

The number of unit bonus payment streams $B^\dagger$ held increases according to the dynamics (7.2.16), where $V^{\star,\dagger}_{Z(t)}(t)$ is the price (on the technical basis) at time t of a unit bonus payment stream. Thus, at time t, the guaranteed payment stream originating from bonus is increased to $s \mapsto Q(t)\,\mathrm{d}B^\dagger(s)$. The technical reserve of the contract, which is formally defined in (7.2.28) below, is the value on the technical basis of the combined benefits given by (7.2.18). By construction, this technical reserve increases exactly by the increase in accumulated dividends, whereas this is typically not the case for the market value; the market value is formally defined in (7.2.26) and (7.2.27) below.

In the remainder of the paper, we focus on valuation of the payment stream (7.2.17), 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.

7.2.5 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:

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

By (7.2.17), 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 (7.2.6) and

$$V^{b}(0) = \mathbb{E}\left[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} Q(t)\,\mathrm{d}B^{\dagger}(t)\right]. \tag{7.2.19}$$

is the time zero market value of bonus payments.

Remark 7.2.1. 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. △

There exist well-established methods to calculate $V^{\circ}(0)$ explicitly using the expected accumulated cash flows of predetermined payments on the market basis from (7.2.8)–(7.2.9); 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 analytical methods and numerical methods for differential equations in regards to 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

$$p^{Q}_{z_0 j}(0,t) = \mathbb{E}\big[Q(t)\mathbb{1}_{(Z(t)=j)} \,\big|\, \mathcal{F}^{S}(t)\big] \tag{7.2.20}$$

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

Proposition 7.2.2. *The time zero market value of bonus payments is given by*

$$V^{b}(0) = \mathbb{E}\left[\int_0^n e^{-\int_0^t r(v)\,\mathrm{d}v} A^{b}(0,\mathrm{d}t)\right], \tag{7.2.21}$$

$$A^{b}(0,\mathrm{d}t) = a^{b}(0,t)\,\mathrm{d}t, \tag{7.2.22}$$

$$a^{b}(0,t) := \sum_{j\in\mathcal{J}} p^{Q}_{z_0 j}(0,t)\Big(b^{\dagger}_{j}(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} b^{\dagger}_{jk}(t)\mu_{jk}(t)\Big). \tag{7.2.23}$$

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

$$p^Q_{z_0 j}(0,t) = Q(t) p_{z_0 j}(0,t), \tag{7.2.24}$$

$$a^b(0,t) = Q(t) a^\dagger(0,t). \tag{7.2.25}$$

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

$$Q(t)\,\mathrm{d}N_{jk}(t) - Q(t)\mathbb{1}_{(Z_{t-}=j)}\mu_{jk}(t)\,\mathrm{d}t$$

defines a martingale, we find that

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

Since there is almost surely at most a finite number of transitions on each compact time interval, we may 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

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

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

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

and thus $a^b(0,t) = Q(t)a^\dagger(0,t)$, cf. (7.2.13). □

Remark 7.2.3. Note that by casting Q according to (7.2.16), by interchanging the order of integration, and by using the law of iterated expectations, it is possible to derive the following alternative formula for the time zero market value of bonus payments:

$$V^b(0) = \int_0^n \mathbb{E}\left[e^{-\int_0^s r(v)\,\mathrm{d}v}\delta(s)\frac{V^\dagger_{Z(s)}(s)}{V^{\star,\dagger}_{Z(s)}(s)}\right]\mathrm{d}s.$$

From this expression, we see how the time zero market value of bonus payments consists of an accumulation of time zero market values of additional benefits bought at different points in time. When we later compare different scenario-based projection models, this representation of the time zero value of the bonus payments turns out to be quite useful, cf. Example 7.4.4. $\triangle$

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 7.3. 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 7.4.

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.

7.2.6 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 (7.2.18). The market value of guaranteed payments V^g is thus given by

$$V^g(t) = \mathbb{E}\left[\int_t^n e^{-\int_t^s r(v)\,\mathrm{d}v} B^g(t,\mathrm{d}s) \,\middle|\, \mathcal{F}(t)\right] = \int_t^n e^{-\int_t^s f(t,v)\,\mathrm{d}v} A^g(t,\,\mathrm{d}s), \quad (7.2.26)$$

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

$$A^g(t,\mathrm{d}s) = A^\circ(t,\mathrm{d}s) + Q(t)A^\dagger_{Z(t)}(t,\mathrm{d}s). \quad (7.2.27)$$

Similarly, the technical reserve of guaranteed payments is given by

$$V^\star(t) = V^{\star,\circ}(t) + Q(t)V^{\star,\dagger}_{Z(t)}(t). \quad (7.2.28)$$

The so-called *portfolio-wide means* of $V^\star$, A^g, 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

$$\begin{aligned}\bar{A}^g(t,s) &:= \mathbb{E}\big[A^g(t,s)\,|\,\mathcal{F}^S(t)\big], \qquad t \le s < \infty\\ \bar{V}^g(t) &:= \mathbb{E}\big[V^g(t)\,|\,\mathcal{F}^S(t)\big],\\ \bar{V}^\star(t) &:= \mathbb{E}\big[V^\star(t)\,|\,\mathcal{F}^S(t)\big]\end{aligned}$$

for $t \ge 0$. From (7.2.26) we find that $\bar{V}^g(t)$ may be obtained from $\bar{A}^g(t,\mathrm{d}s)$ via the equation

$$\bar{V}^g(t) = \int_t^n e^{-\int_t^s f(t,v)\,\mathrm{d}v}\,\bar{A}^g(t,\mathrm{d}s). \tag{7.2.29}$$

Thus it suffices to consider only $\bar{A}^g$ and $\bar{V}^\star$.

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 7.3.1, 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 $U = \{U(t)\}_{t\ge 0}$. 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 $(\theta,\eta) = (\theta(t),\eta(t))_{t\ge 0}$ 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

$$\mathrm{d}U(t) = r(t)(U(t) - \eta(t)S_1(t))\mathrm{d}t + \eta(t)\,\mathrm{d}S_1(t) - \big(A^\circ(0,\mathrm{d}t) + A^b(0,\mathrm{d}t)\big). \tag{7.2.30}$$

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 take into account multiple independent insured, 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

$$\sum_{j \in \mathcal{J}} w_j \mathbb{E}_j\big[\, A^g(t,s) \,|\, \mathcal{F}^S(t)\big] \qquad \text{and} \qquad \sum_{j \in \mathcal{J}} w_j \mathbb{E}_j\big[\, V^\star(t) \,|\, \mathcal{F}^S(t)\big],$$

where $\mathbb{E}_j$ 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 $S(\cdot \wedge t) = \{S(u)\}_{0 \leq u \leq t}$. We can now make the concepts of shape and controls precise.

Definition 7.2.4. *The shape of the insurance business $\mathcal{I}$ is the triplet*

$$\mathcal{I} = \big\{U(t), \bar{A}^g(t, \mathrm{d}s), \bar{V}^\star(t)\big\}_{t \geq 0},$$

while the controls are the pair $(\delta, \eta) = \{\delta(t), \eta(t)\}_{t \geq 0}$.

Assumption 7.2.5. *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*

$$\eta(t) = \eta(t, S(\cdot \wedge t), \mathcal{I}(t)) \tag{7.2.31}$$

for some explicitly computable and suitably regular deterministic mapping η, and we assume that δ takes the form

$$\begin{aligned} \delta(t) = {} & \delta_0\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) \\ & + \delta_1\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) \rho(\tau)^{\mathbb{1}_{(\tau \leq t)}} \\ & + \delta_2\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) Q(t), \end{aligned} \tag{7.2.32}$$

for some suitably regular deterministic mappings δ_0, δ_1 and δ_2 that we are able to compute explicitly.

Remark 7.2.6. In Remark 7.2.1 we discussed the extension to general $Q(0)$ and the idea of focusing on $Q(\cdot) - Q(0)$. By rewriting (7.2.32) in the following manner,

$$\begin{aligned} \delta(t) = {} & \delta_0\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) + \delta_2\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) Q(0) \\ & + \delta_1\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) \rho(\tau)^{\mathbb{1}_{(\tau \leq t)}} \\ & + \delta_2\left(t, S(\cdot \wedge t), Z(t), \mathcal{I}(t)\right) (Q(t) - Q(0)), \end{aligned}$$

we see how this idea would manifest itself in relation to Assumption 7.2.5. $\triangle$

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. (7.2.17). This is important for practical applications, as the following example highlights.

Example 7.2.7 (Second order interest rate)**.** The technical reserve $V^\star$ from (7.2.28) accumulates with the first order interest rate $r^\star$. Dividends may then arise by accumulating the technical reserve with a second order interest rate r^δ that is continuously readjusted based on the shape of the insurance business. This is obtained by letting

$$\delta(t) = \big(r^\delta(t) - r^\star(t)\big)V^\star(t), \tag{7.2.33}$$

$$r^\delta(t) = \Phi(t, S(\cdot \wedge t), \mathcal{I}(t)), \tag{7.2.34}$$

for some explicitly computable and suitably regular mapping Φ. This corresponds to setting

$$\begin{aligned}
\delta_0(t,j) &= \big(r^\delta(t) - r^\star(t)\big)1_{(j \in \mathcal{J}^p \setminus \{J\})}\widetilde{V}^\star_j(t),\\
\delta_1(t,j) &= \big(r^\delta(t) - r^\star(t)\big)1_{(j \in \mathcal{J}^f \setminus \{2J+1\})}\widetilde{V}^{\star,+}_{j'}(t)\\
\delta_2(t,j) &= \big(r^\delta(t) - r^\star(t)\big)V^{\star,\dagger}_j(t),
\end{aligned}$$

for all $j \in \mathcal{J}$. In case of a survival model, various specifications of second order interest rates is among the focal points of Møller and Steffensen (2007). In discrete time, second order interest rates are also employed in Bacinello (2001), where r^δ is taken to be the maximum of the technical interest rate $r^\star$ and a proportion (the so-called participation level) of the return on some reference portfolio. In the numerical example of Section 7.5, we consider the following special case of r^δ:

$$r^\delta(t) = r^\star(t) + \frac{\kappa\big(U(t) - \max\big\{\bar{V}^\star(t), \bar{V}^g(t)\big\}\big)^+}{\bar{V}^\star(t)}, \quad \kappa \in [0,1], \tag{7.2.35}$$

where κ is the share of some measure of excess assets used to buy additional benefits. This choice of r^δ, where $r^\delta > r^\star$, in particular leads to a dividend design which also constitutes a minimum interest rate guarantee of $r^\star$. In case of a survival model, this type of r^δ also appears in e.g. Møller and Steffensen (2007, Chapter 4.5.3). ◦

The aim of this paper is to develop methods to compute the market value of bonus payments $V^b(0)$. Recall from Proposition 7.2.2 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 7.2.5, we adopt a simulation approach. It follows from (7.2.16) that for a simulated financial scenario, i.e. a realization of the whole path of S, we need the shape of the insurance business $\mathcal{I}(t) = (U(t), \bar{A}^g(t, \mathrm{d}s), \bar{V}^\star(t))$ 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 7.3 concerns the general case where Q is allowed to be $\mathcal{F}^Z \vee \mathcal{F}^S$-adapted and where we apply (7.2.22)–(7.2.23). In the subsequent Section 7.4 we specialize to Q being state independent (of Z), i.e. $\mathcal{F}^S$-adapted, where we instead can apply the simpler formula (7.2.25).

7.3 Scenario-based projection model

This section contains the main contributions of the paper and provides the foundation for the special case in Section 7.4. In Subsection 7.3.1, 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 7.3.2 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 7.3.3 a procedure for the computation of $V^b(0)$ via an application of Proposition 7.2.2.

As noted in Proposition 7.2.2, 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 7.4.

7.3.1 Projecting the shape

We now turn our attention to projection of the shape of the insurance business. This consists of computation of $\mathcal{I} = (U, \bar{A}^g, \bar{V}^\star)$ 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 (7.2.30). The computational issue reduces to that of computing $p^Q_{z_0 j}(0,\cdot)$, cf. (7.2.22)–(7.2.23) and (7.2.30). Thus we focus on the projection of the portfolio-wide means $\bar{A}^g$ and $\bar{V}^\star$.

First, we consider the portfolio-wide means $\bar{A}^g$ of the expected accumulated guaranteed cash flows A^g.

Proposition 7.3.1. *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 (7.2.27), (7.2.20), and due to the assumed independence between Z and S, we immediately find that

$$\begin{aligned}\bar{A}^g(t,s) &= \mathbb{E}\big[A^\circ(t,s)\,|\,\mathcal{F}^S(t)\big] + \sum_{j\in\mathcal{J}} \mathbb{E}\Big[\mathbb{1}_{(Z(t)=j)} Q(t) A^\dagger_{Z(t)}(t,s)\,\Big|\,\mathcal{F}^S(t)\Big] \\ &= \mathbb{E}[A^\circ(t,s)] + \sum_{j\in\mathcal{J}} p^Q_{z_0 j}(0,t) A^\dagger_j(t,s).\end{aligned}$$

By (7.2.7) and the iterated law of expectations,

$$\begin{aligned}\mathbb{E}[A^\circ(t,s)] &= \mathbb{E}[B^\circ(s) - B^\circ(t)] \\ &= A^\circ(0,s) - \mathbb{E}[B^\circ(t) - B^\circ(0)].\end{aligned}$$

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. □

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 (7.2.4), cf. Subsection 7.2.3 and Subsection 7.2.4.

Denote by $\bar{V}^{\star,\circ}$ the portfolio-wide mean technical reserves of predetermined payments given by

$$\bar{V}^{\star,\circ}(t) = \mathbb{E}\big[V^{\star,\circ}(t)\,|\,\mathcal{F}^S(t)\big]$$

for $t \geq 0$. Since Z and S are assumed independent, we could replace the conditional expectation by an ordinary expectation.

Proposition 7.3.2. *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

$$\bar{V}^{\star,\circ}(t) = \sum_{\substack{j \in \mathcal{J}^p \\ j \neq J}} p_{z_0 j}(0,t) \widetilde{V}^\star_j(t) + \sum_{\substack{j \in \mathcal{J}^f \\ j \neq 2J+1}} p^\rho_{z_0 j}(0,t) \widetilde{V}^{\star,+}_{j'}(t). \tag{7.3.1}$$

Proof. By (7.2.28) and (7.2.20), direct calculations yield

$$\begin{aligned}
\bar{V}^\star(t) &= \mathbb{E}\big[V^{\star,\circ}(t) \,|\, \mathcal{F}^S(t)\big] + \sum_{j \in \mathcal{J}} \mathbb{E}\Big[\mathbb{1}_{(Z(t)=j)} Q(t) V^{\star,\dagger}_{Z(t)}(t) \,\Big|\, \mathcal{F}^S(t)\Big] \\
&= \bar{V}^{\star,\circ}(t) + \sum_{j \in \mathcal{J}} p^Q_{z_0 j}(0,t) V^{\star,\dagger}_j(t).
\end{aligned}$$

To obtain (7.3.1), we split $V^{\star,\circ}$ according to the events of $Z(t)$ being in $\mathcal{J}^p \setminus \{J\}$, $\mathcal{J}^f \setminus \{2J+1\}$, and $\{J, 2J+1\}$. According to (7.2.5), we then have

$$\begin{aligned}
\bar{V}^{\star,\circ}(t) &= \mathbb{E}\Big[\mathbb{1}_{(Z(t)\in\mathcal{J}^p\setminus\{J\})} \widetilde{V}^\star_{Z(t)}(t) + \mathbb{1}_{(Z(t)\in\mathcal{J}^f\setminus\{2J+1\})} \rho(\tau) \widetilde{V}^{\star,+}_{Z(t)'}(t) \,\Big|\, \mathcal{F}^S(t)\Big] \\
&= \mathbb{E}\Bigg[\sum_{\substack{j \in \mathcal{J}^p \\ j \neq J}} \mathbb{1}_{(Z(t)=j)} \widetilde{V}^\star_j(t) + \sum_{\substack{j \in \mathcal{J}^f \\ j \neq 2J+1}} \mathbb{1}_{(Z(t)=j)} \rho(\tau) \widetilde{V}^{\star,+}_{j'}(t) \,\Bigg|\, \mathcal{F}^S(t)\Bigg] \\
&= \sum_{\substack{j \in \mathcal{J}^p \\ j \neq J}} p_{z_0 j}(0,t) \widetilde{V}^\star_j(t) + \sum_{\substack{j \in \mathcal{J}^f \\ j \neq 2J+1}} p^\rho_{z_0 j}(0,t) \widetilde{V}^{\star,+}_{j'}(t),
\end{aligned}$$

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 (7.2.10). Thus Proposition 7.3.2 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.

7.3.2 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 $z_0 \in \mathcal{J}^f$, which is in accordance with $\tau = 0$ for $z_0 \in \mathcal{J}^f$ and the assumption $\rho(0) = 1$.

Theorem 7.3.3. *The Q-modified transition probabilities $p^Q_{z_0 j}(0,\cdot)$ satisfy for $j \in \mathcal{J}$ the differential equations*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} p^Q_{z_0 j}(0,t) = {} & \frac{p_{z_0 j}(0,t)\delta_0(t,j) + p^\rho_{z_0 j}(0,t)\delta_1(t,j) + p^Q_{z_0 j}(0,t)\delta_2(t,j)}{V^{\star,\dagger}_j(t)} \\ & - p^Q_{z_0 j}(0,t)\mu_{j\bullet}(t) + \sum_{\substack{k\in\mathcal{J}\\ k\neq j}} p^Q_{z_0 k}(0,t)\mu_{kj}(t), \qquad p^Q_{z_0 j}(0,0) = 0.\end{aligned} \tag{7.3.2}$$

Proof. The boundary conditions follows by the assumption that $Q(0) = 0$. Referring to (7.2.20) and (7.2.16), we have

$$p^Q_{z_0 j}(0,t) = \mathbb{E}\big[\mathbb{1}_{(Z(t)=j)} Q(t) \,\big|\, \mathcal{F}^S(t)\big] = \mathbb{E}\left[\mathbb{1}_{(Z(t)=j)} \int_0^t \frac{\delta(u)}{V^{\star,\dagger}_{Z(u)}(u)} \,\mathrm{d}u \,\middle|\, \mathcal{F}^S(t)\right]$$

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 $k \in \mathcal{J}$,

$$\begin{aligned}\mathbb{E}\left[\mathbb{1}_{(Z(u)=k)} \frac{p^Q_{z_0 k}(0,u)}{p_{z_0 k}(0,u)} \,\middle|\, \mathcal{F}^S(t)\right] &= \mathbb{E}\big[\mathbb{1}_{(Z(u)=k)} Q(u) \,\big|\, \mathcal{F}^S(t)\big], \\ \mathbb{E}\left[\mathbb{1}_{(Z(u)=k)} \frac{p^\rho_{z_0 k}(0,u)}{p_{z_0 k}(0,u)}\right] &= \mathbb{E}\big[\mathbb{1}_{(Z(u)=k)} \rho(\tau)^{\mathbb{1}_{(\tau\leq u)}}\big].\end{aligned}$$

Thus by Markovianity of Z and independence between Z and S,

$$p^Q_{z_0 j}(0,t) = \mathbb{E}\left[\mathbb{1}_{(Z(t)=j)} \int_0^t \sum_{k\in\mathcal{J}} \mathbb{1}_{(Z(u)=k)} b^Q_k(u) \,\mathrm{d}u \,\middle|\, \mathcal{F}^S(t)\right] \tag{7.3.3}$$

with b^Q_k, $k \in \mathcal{J}$, given by

$$b^Q_k(u) = \frac{\delta_0(u,k) + \delta_1(u,k)\frac{p^\rho_{z_0 k}(0,u)}{p_{z_0 k}(0,u)} + \delta_2(u,k)\frac{p^Q_{z_0 k}(0,u)}{p_{z_0 k}(0,u)}}{V^{\star,\dagger}_k(u)} \tag{7.3.4}$$

for all $u \geq 0$. The assumption of independence between Z and S, Markovianity of Z, and Fubini's theorem finally yield

$$p^Q_{z_0 j}(0,t) = \int_0^t \sum_{k\in\mathcal{J}} p_{z_0 k}(0,u) p_{kj}(u,t) b^Q_k(u) \,\mathrm{d}u. \tag{7.3.5}$$

The statement of the theorem is now established by differentiation as follows. Leibniz' integration rule gives

$$\frac{\mathrm{d}}{\mathrm{d}t} p^Q_{z_0 j}(0,t) = \sum_{k\in\mathcal{J}} 1_{(k=j)} p_{z_0 k}(0,t) b^Q_k(t)$$

$$
\begin{aligned}
&+ \int_0^t \sum_{k \in \mathcal{J}} p_{z_0 k}(0,u) \left(\frac{\mathrm{d}}{\mathrm{d}t} p_{kj}(u,t) \right) b_k^Q(u) \, \mathrm{d}u \\
&= \frac{\delta_0(t,j) p_{z_0 j}(0,t) + \delta_1(t,j) p_{z_0 j}^{\rho}(0,t) + \delta_2(t,j) p_{z_0 j}^{Q}(0,t)}{V_j^{\star,\dagger}(t)} \\
&+ \int_0^t \sum_{k \in \mathcal{J}} p_{z_0 k}(0,u) \left(\frac{\mathrm{d}}{\mathrm{d}t} p_{kj}(u,t) \right) b_k^Q(u) \, \mathrm{d}u.
\end{aligned}
$$

Applying Kolmogorov's forward differential equations and (7.3.5) to the last line of the equation we find that

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p_{z_0 j}^{Q}(0,t) = {} & \frac{\delta_0(t,j) p_{z_0 j}(0,t) + \delta_1(t,j) p_{z_0 j}^{\rho}(0,t) + \delta_2(t,j) p_{z_0 j}^{Q}(0,t)}{V_j^{\star,\dagger}(t)} \\
& - p_{z_0 j}^{Q}(0,t) \mu_{j\bullet}(t) + \sum_{\substack{\ell \in \mathcal{J} \\ \ell \neq j}} p_{z_0 \ell}^{Q}(0,t) \mu_{\ell j}(t)
\end{aligned}
$$

as desired. □

Remark 7.3.4. There exists a clear link between Q-modified transition probabilities and so-called state-wise retrospective reserves. Referring to (7.3.3) and (7.3.4), we see that for a fixed financial scenario,

$$
W_j(\cdot) := \frac{p_{z_0 j}^{Q}(0,\cdot)}{p_{z_0 j}(0,\cdot)}
$$

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. Subsection 5.B in Norberg, 1991) with payments

$$
- \sum_{j \in \mathcal{J}} \mathbb{1}_{(Z(t)=j)} b_j^Q(t) \ \mathrm{d}t
$$

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 $p_{z_0 j}^{Q}(0,\cdot)$ from Theorem 7.3.3 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 7.3.3 allows us to formulate a procedure for the calculation of $V^b(0)$. The procedure is presented in the next subsection.

7.3.3 Numerical procedure

Based on the results of the previous subsections, we demonstrate a procedure for the scenario-based projection model. An actual numerical example is given later in Section 7.5. In what follows, we suppose we are given mappings (δ, η) serving as controls. They are assumed to satisfy Assumption 7.2.5.

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 (7.2.9).

(2) The portfolio-wide mean technical reserve of predetermined payments $\bar{V}^{\star,\circ}(t)$ for all $t \geq 0$ calculated via (7.3.1).

(3) For each $t \geq 0$, state-wise expected accumulated unit bonus cash flows $A_j^{\dagger}(t, s)$ for all $s \geq t$ and $j \in \mathcal{J}$ as in (7.2.14)–(7.2.15).

(4) State-wise technical unit reserves $V_j^{\star,\dagger}(t)$ for all $t \geq 0$ and $j \in \mathcal{J}$ as in (7.2.11).

(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 7.3.3 and (7.2.30), respectively.
  - Evaluating the differentials at time t involves the mappings $(\delta_0, \delta_1, \delta_2, \eta)$ from (7.2.31)–(7.2.32). By inspection of the differentials and these mappings, we see that we require the shape of the insurance business

$$\mathcal{I}^k(t) = \big(U^k(t), \bar{A}^{g,k}(t, \mathrm{d}s), \bar{V}^{\star,k}(t)\big),$$

the expected bonus cash flow $a^{b,k}(0,t)$, as well as the input. Computation of $\bar{A}^{g,k}(t,\mathrm{d}s)$, $\bar{V}^{\star,k}(t)$, and $a^{b,k}(0,t)$ is achieved via Proposition 7.3.1, Proposition 7.3.2, and (7.2.23).

- 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 $A_j^{\dagger}(\cdot,\cdot)$ 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 general, it is preferable to use analytical methods and numerical methods for ordinary differential equations compared to Monte Carlo methods. For example, solving Kolmogorov's forward differential equations in order to calculate an expected cash flow may be done orders of magnitudes faster (for a given precision requirement) compared to calculating the same expected cash flow via Monte Carlo methods. The procedure presented above shows how to disentangle biometric and behavioral risk from financial risk, allowing us to solve part of the problem via numerical methods for differential equations instead of using Monte Carlo methods. For a given precision requirement, this significantly reduces the time required to handle biometric and behavioral risk. If the portfolio merely consists of a single insured, which we presumed, it is however the simulation of financial risk that constitutes the main numerical complexity and time usage. As the number of insured increases, so does the relative time needed to calculate expected cash flows, since the same financial scenarios are used across all insured, and thus the numerical speed up of our procedure, compared with a full-blown Monte Carlo approach, should become significant.

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.

7.4 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 7.3 which relies on ensuring Q to be an $\mathcal{F}^S$-adapted process such that the simplified case of Proposition 7.2.2 applies.

In Subsection 7.4.2, we provide sufficient conditions on δ such that Q is $\mathcal{F}^S$-adapted. Next, Subsection 7.4.3 revisits the projection of the shape under this simplification. Finally, in Subsection 7.4.4 we present a procedure for the computation of the market value of bonus payments in the state-independent projection model.

7.4.1 Background

The concept of state-independent modeling is uncommon in the literature, with the projection model described in Jensen and Schomacker (2015, Section 4) being one of few exceptions. It is our impression that projections models such as the one found in Jensen and Schomacker (2015, Section 4) have been implemented or are being implemented in practice, which further underlines the importance of studying state-independent scenario-based projection models in more detail.

In the projection model described in Jensen and Schomacker (2015, Section 4), additional benefits are bought according to the portfolio-wide mean $\bar{V}^{\star,\dagger}$ of the technical reserve rather than the actual technical reserve $V^{\star,\dagger}_{Z(\cdot)}$, see Jensen and Schomacker (2015, p. 196). Furthermore, the dividend yield is implicitly assumed $\mathcal{F}^S$-adapted, since in Jensen and Schomacker (2015) the evolution of the policy is only described on an averaged portfolio level. In unity, this leads to an $\mathcal{F}^S$-adapted Q. In the following, we provide sufficient conditions on δ within our setup to ensure this adaptability. Later, in Example 7.4.4 in Subsection 7.4.3, we provide a slightly more explicit link to the projection model in Jensen and Schomacker (2015, Section 4).

7.4.2 Class of dividend strategies

Recall from (7.2.16) and (7.2.32) 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

$$\delta_i(t, Z(t)) = \widetilde{\delta}_i(t) V^{\star,\dagger}_{Z(t)}(t), \qquad i = 0, 2, \tag{7.4.1}$$

$$\delta_1(t, Z(t)) = 0, \tag{7.4.2}$$

where we have used the shorthand notation $\widetilde{\delta}_i(t) = \widetilde{\delta}_i\,(t, S(\cdot \wedge t), \mathcal{I}(t))$ for suitably regular deterministic mappings $\widetilde{\delta}_i$, $i = 0, 2$. This is a consequence of the following observation. When (7.4.1)–(7.4.2) hold, then simply

$$\mathrm{d}Q(t) = \big(\widetilde{\delta}_0(t) + \widetilde{\delta}_2(t) Q(t)\big)\mathrm{d}t, \quad Q(0) = 0. \tag{7.4.3}$$

This implies $p^Q_{z_0 j}(0, t) = Q(t) p_{z_0 j}(0, t)$, cf. (7.2.24).

Remark 7.4.1. Since the class of dividend strategies presented here builds on Assumption 7.2.5, affinity in Q is more or less implicitly assumed. The simplifications we obtain in the following Subsections 7.4.3–7.4.4 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. △

7.4.3 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 7.4.2. *Assume that the dividend strategy δ is on the form* (7.4.1)–(7.4.2). *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 7.3.1 and its proof, we have

$$\bar{A}^g(t, s) = A^\circ(0, s) - E[B^\circ(t) - B^\circ(0)] + \mathbb{E}\big[\, Q(t) A^\dagger(t, s) \,\big|\, \mathcal{F}^S(t)\big]\,.$$

Since by assumption Q is $\mathcal{F}^S$-adapted and Z and S are independent, referring to (7.2.6) with superscript $\circ$ replaced by $\dagger$ and applying the law of iterated expectations yields

$$\begin{aligned}
\mathbb{E}\big[\, Q(t) A^\dagger(t, s) \,\big|\, \mathcal{F}^S(t)\big] &= Q(t) \mathbb{E}\big[B^\dagger(s) - B^\dagger(t)\big] \\
&= Q(t) A^\dagger(0, s) - Q(t) \mathbb{E}\big[B^\dagger(t) - B^\dagger(0)\big]
\end{aligned}$$

Consequently,

$$\bar{A}^g(t,\, \mathrm{d}s) = A^\circ(0,\, \mathrm{d}s) + Q(t) A^\dagger(0,\, \mathrm{d}s)$$

as desired. □

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

$$\bar{V}^{\star,\dagger}(t) = \mathbb{E}\Big[V^{\star,\dagger}_{Z(t)}(t) \,\Big|\, \mathcal{F}^S(t) \Big]$$

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

$$\bar{V}^{\star,\dagger}(t) = \sum_{j \in \mathcal{J}} p_{z_0 j}(0,t) V^{\star,\dagger}_j(t). \tag{7.4.4}$$

Corollary 7.4.3. *Assume that the dividend strategy δ is on the form* (7.4.1)–(7.4.2)*. 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 (7.2.24), Proposition 7.3.2, and (7.4.4). □

The following example is a continuation of Example 7.2.7 regarding the accumulation of the technical reserve with a second order interest rate.

Example 7.4.4 (Second order interest rate continued)**.** The dividend strategy from Example 7.2.7 regarding accumulation of the technical reserve $V^\star$ with a second order interest rate r^δ does not satisfy the requirements on δ from (7.4.1)–(7.4.2). Instead, the strategy

$$\delta(t) = \big(r^\delta(t) - r^\star(t)\big)\frac{\bar{V}^\star(t)}{\bar{V}^{\star,\dagger}(t)} V^{\star,\dagger}_{Z(t)}(t), \tag{7.4.5}$$

satisfies (7.4.1)–(7.4.2) 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 (7.4.3),

$$\bar{V}^{\star,\dagger}(t)\,\mathrm{d}Q(t) = \big(r^\delta(t) - r^\star(t)\big)\bar{V}^\star(t)\,\mathrm{d}t.$$

This is in accordance with the projection model in Jensen and Schomacker (2015, Section 4). By multiplying the strategy (7.4.5) 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 7.2.7. If the two ratios are close to one, the strategy (7.4.5) approximates the strategy of Example 7.2.7. Note that

$\mathbb{E}\big[V^\star(t)/\bar{V}^\star(t) \,\big|\, \mathcal{F}^S(t)\big] = 1$, 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 $V^{\star,\dagger}_{Z(t)}(t)$.

Even if the ratioes are not close to one, the strategy given by (7.4.5) may still approximate the strategy of Example 7.2.7 in terms of producing an akin time zero market value of bonus payments. Taking as starting point the strategy of Example 7.2.7 and the expression of Remark 7.2.3, we find that

$$V^b(0) = \int_0^n \mathbb{E}\left[e^{-\int_0^s r(v)\,\mathrm{d}v}\big(r^\delta(s) - r^\star(s)\big)V^\star(s)\frac{V^\dagger_{Z(s)}(s)}{V^{\star,\dagger}_{Z(s)}(s)}\right] \mathrm{d}s,$$

so that by the law of iterated expectations,

$$V^b(0) = \int_0^n \mathbb{E}\left[e^{-\int_0^s r(v)\,\mathrm{d}v}\big(r^\delta(s) - r^\star(s)\big)\mathbb{E}\left[V^\star(s)\frac{V^\dagger_{Z(s)}(s)}{V^{\star,\dagger}_{Z(s)}(s)} \,\middle|\, \mathcal{F}^S(s)\right]\right] \mathrm{d}s. \tag{7.4.6}$$

Taking instead as a starting point the strategy given by (7.4.5), we find the expression

$$\int_0^n \mathbb{E}\left[e^{-\int_0^s r(v)\,\mathrm{d}v}\big(r^\delta(s) - r^\star(s)\big)\bar{V}^\star(s)\frac{V^\dagger_{Z(s)}(s)}{\bar{V}^{\star,\dagger}(s)}\right] \mathrm{d}s \tag{7.4.7}$$

for the corresponding time zero market value of bonus payments. We should like to hightlight that the second order interest rate r^δ depends on Q through the shape of the insurance business, so that the second order interest rates of (7.4.6) and (7.4.7) may differ.

From (7.4.6) and (7.4.7), we see that the strategy given by (7.4.5) in particular leads to a decent approximation of the market value of bonus payments as long as

$$\begin{aligned}
&\big(r^\delta(s) - r^\star(s)\big)\mathbb{E}\left[V^\star(s)\frac{V^\dagger_{Z(s)}(s)}{V^{\star,\dagger}_{Z(s)}(s)} \,\middle|\, \mathcal{F}^S(s)\right] \\
&\qquad\approx \big(r^\delta(s) - r^\star(s)\big)\bar{V}^\star(s)\frac{\mathbb{E}\Big[V^\dagger_{Z(s)}(s) \,\Big|\, \mathcal{F}^S(s)\Big]}{\bar{V}^{\star,\dagger}(s)}
\end{aligned}$$

for all $s \geq 0$. This is for example the case if the second order interest rates only differ ever so slightly and

$$\mathbb{E}\left[V^\star(s)\frac{V^\dagger_{Z(s)}(s)}{V^{\star,\dagger}_{Z(s)}(s)} \,\middle|\, \mathcal{F}^S(s)\right] \approx \bar{V}^\star(s)\frac{\mathbb{E}\Big[V^\dagger_{Z(s)}(s) \,\Big|\, \mathcal{F}^S(s)\Big]}{\bar{V}^{\star,\dagger}(s)}$$

for all $s \geq 0$. The latter is by definition of the portfolio-wide mean technical reserve $\bar{V}^\star$ indeed the case if the safety loading $s \mapsto V^\dagger_{Z(s)}(s)/V^{\star,\dagger}_{Z(s)}(s)$ is approximately state-independent, i.e. does not depend significantly on Z (whenever $\bar{V}^\star$ is non-zero). ◦

7.4.4 Numerical procedure

Based on the results of the previous subsections, we demonstrate a procedure for the state-independent scenario-based projection model. An actual numerical example is given later in Section 7.5. In what follows, we suppose we are given mappings (δ, η) serving as controls. They are assumed to satisfy Assumption 7.2.5 with δ on the form (7.4.1)–(7.4.2).

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 (7.2.9).

(2) The portfolio-wide mean technical reserve of predetermined payments $\bar{V}^{\star,\circ}(t)$ for all $t \geq 0$ calculated via (7.3.1).

(3) The expected unit bonus cash flow $a^{\dagger}(0, s)$ for all $s \geq 0$ as in (7.2.13).

(4) The portfolio-wide mean technical unit bonus reserve $\bar{V}^{\star,\dagger}(t)$ for all $t \geq 0$ calculated via (7.4.4)

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 (7.4.3) and (7.2.30), respectively.
  - Evaluating the differentials at time t involves the mappings $(\widetilde{\delta}_0, \widetilde{\delta}_2, \eta)$ from (7.2.31) and (7.4.1). By inspection of the differentials and these mappings, we see that we require the shape of the insurance business

$$\mathcal{I}^k(t) = \big(U^k(t), \bar{A}^{g,k}(t, \mathrm{d}s), \bar{V}^{\star,k}(t)\big),$$

the expected bonus cash flow $a^{b,k}(0,t) = Q^k(t)a^\dagger(0,t)$, cf. (7.2.25), as well as the input. Computation of $\bar{A}^{g,k}(t, \mathrm{d}s)$ and $\bar{V}^{\star,k}(t)$ is achieved via Corollary 7.4.2 and Corollary 7.4.3.

- 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 7.3.3, 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 (7.4.1)–(7.4.2).

7.5 Numerical example

In this section, we illustrate the methods presented in the previous sections via a numerical example intended to show how our methods and results can be applied in practice. The predetermined payments, technical basis, and market basis are based on the numerical example in Buchardt and Møller (2015); our extension consists of the inclusion of financial risk and bonus payments. The numerical example aims at illustrating similarities and differences between the state-dependent scenario-based projection model and a state-independent approximation in the spirit of Jensen and Schomacker (2015, Section 4), cf. Example 7.4.4.

7.5.1 Setup

The state of the insured is modeled in an eight-state disability model with policyholder behavior as depicted in Figure 7.2. We consider a male who is 40 years old today. His retirement age is taken to be 65, and his predetermined payments consist of:

- A disability annuity of rate 100000 per year while disabled, but only until retirement, i.e. age 65.
- A life annuity of rate 100000 per year while alive and non-lapsed (corresponding to states 0, 1, 4, and 5), but only from retirement, i.e. age 65.

- Premium payments of rate 46409.96 per year while active, but only until retirement, i.e. age 65.

The maximal contract time is set equal to 70, i.e. $n = 70$, corresponding to a maximal age of $40 + 70 = 110$ years. We note that the predetermined payments

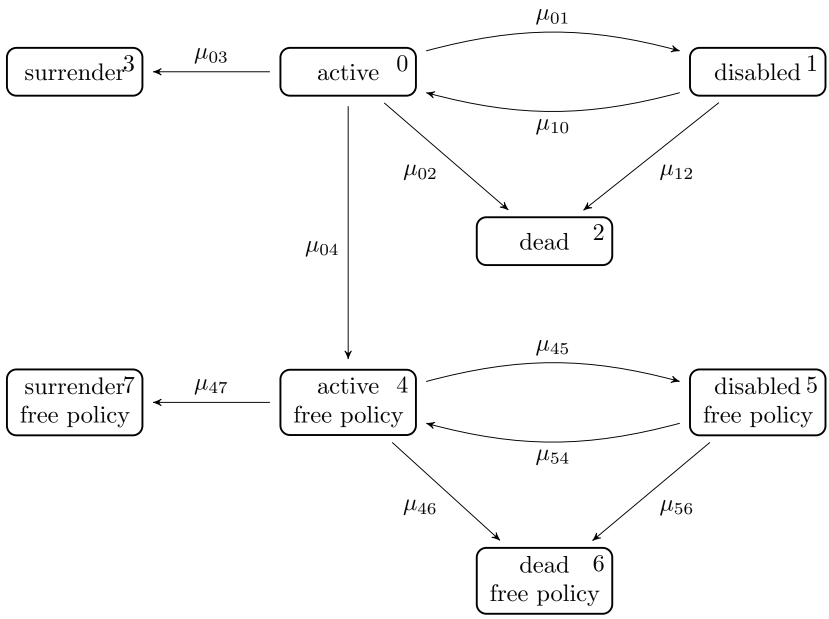

Figure 7.2: *Disability model with policyholder behavior. The state-space is decomposed according to $\mathcal{J} = \mathcal{J}^p \cup \mathcal{J}^f$ with $\mathcal{J}^p = \{0, \ldots, 3\}$ and $\mathcal{J}^f = \{4, \ldots, 7\}$.*

are actuarially fair in the sense that the equivalence principle is satisfied on the technical basis. The technical basis takes the following form:

$$
\begin{aligned}
r^\star(s) &= 0.01, \\
\mu^\star_{01}(s) &= \Big(0.0004 + 10^{4.54+0.06(s+40)-10}\Big)\mathbb{1}_{(s\leq 25)}, \\
\mu^\star_{10}(s) &= \Big(2.0058e^{-0.117(s+40)}\Big)\mathbb{1}_{(s\leq 25)}, \\
\mu^\star_{02}(s) &= 0.0005 + 10^{5.88+0.038(s+40)-10}, \\
\mu^\star_{12}(s) &= \mu^\star_{02}(s)\big(1 + \mathbb{1}_{(s\leq 25)}\big).
\end{aligned}
$$

The technical basis and the aforementioned predetermined payments further determine the surrender payments and the free policy factor, which we do not explicitly state, cf. Subsection 7.2.3.

The market basis takes the following form:

$\mu_{02}(\cdot)$: The 2012 edition of the Danish FSA's longevity benchmark,

$$\begin{aligned}
\mu_{01}(s) &= 10^{5.662015+0.033462(s+40)-10}\mathbb{1}_{(s\le 25)},\\
\mu_{10}(s) &= 4.0116e^{-0.117(s+40)}\mathbb{1}_{(s\le 25)},\\
\mu_{12}(s) &= \left(0.010339+10^{5.070927+0.05049(s+40)-10}\right)\mathbb{1}_{(s\le 25)}+\mu_{02}(s)\mathbb{1}_{(s>25)},\\
\mu_{03}(s) &= (0.06-0.002s)\mathbb{1}_{(s\le 25)},\\
\mu_{04}(s) &= 0.05\mathbb{1}_{(s\le 25)},
\end{aligned}$$

with $\mu_{jk}=\mu_{(j-4)(k-4)}$ for $j,k\in\mathcal{J}^f$, $j\neq k$.

We now deviate from Buchardt and Møller (2015) by introducing a bond market, so that the risky asset with price process S_1 corresponds to a zero-coupon bond with expiry $n=70$. The short rate follows a Vasicek model, so that the dynamics are given by

$$\begin{aligned}
\mathrm{d}r(t) &= (\beta-\alpha\cdot r(t))\mathrm{d}t+\sigma\mathrm{d}W(t), \quad r(0)=0.01,\\
\mathrm{d}S_1(t) &= r(t)S_1(t)\mathrm{d}t-\sigma\psi(t,n)S_1(t)\mathrm{d}W(t), \quad S_1(n)=1,
\end{aligned}$$

where $\psi(t,n)=\left(1-e^{-\alpha(n-t)}\right)/\alpha$. The parameters β, α, and σ may be found in Table 7.1. They are taken from the numerical example in Buchardt and Møller (2018) and yield a mean reversion to about 0.043.

Regarding bonus payments, we consider the case where all dividends are used only to buy additional life annuity benefits, so that in particular the rate of the disability annuity is kept fixed throughout the entire contract period. Consequently, the unit bonus payments $B^\dagger$ are determined according to

$$b_0^\dagger(t)=b_1^\dagger(t)=100000\cdot\mathbb{1}_{(t\ge 25)}.$$

Regarding the controls (δ,η), we consider the dividend strategy δ introduced in Example 7.2.7 with the second order interest rate (7.2.35), and investment strategy η given by:

$$\eta(t)=\frac{\int_t^n\psi(t,s)e^{-\int_t^s f(t,v)\,\mathrm{d}v}\bar{A}^g(t,\mathrm{d}s)}{\psi(t,n)S_1(t)}.$$

The investment strategy is chosen such that it hedges the interest rate risk of the guaranteed cash flows $\bar{A}^g$ on a portfolio level.

The values of the parameters for the short rate model and the dividend strategy are shown in Table 7.1.

Parameter	β	α	σ	κ
Value	0.007006001	0.162953	0.015384	0.2

Table 7.1: *Parameters for the short rate model and dividend strategy. The parameters of the former are taken from the numerical example in Buchardt and Møller (2018), which provides a mean reversion of the short rate to* $\beta/\alpha\approx 0.043$.

7.5.2 Results and discussion

The inputs as described in Subsection 7.3.3 and Subsection 7.4.4 are computed in classic fashion using standard numerical methods. Next, we carry out the state-dependent numerical procedure outlined in Subsection 7.3.3 as well as the state-independent procedure presented in Subsection 7.4.4 to determine the time zero market value of bonus payments $V^b(0)$; the computations are based $N = 10000$ financial scenarios and Euler-Maruyama discretizations with step length 0.01 years. In the latter procedure, we use the dividend strategy presented in Example 7.4.4 with r^δ on the same form as above. The results are presented in Table 7.2 along with the market value of predetermined payments $V^\circ(0)$.

$V^\circ(0)$	Time zero market value of bonus payments $V^b(0)$		
	State-dependent	State-independent	Relative difference
-72582	72661	72663	-0.00201%

Table 7.2: *Time zero market values from both the state-dependent and state-independent implementation. The relative difference lies within the margin of numerical error.*

We see that the two implementations produce identical results, in the sense that the difference is within the margin of numerical error, for this product design and set of parameter values.

To show what is going on behind the scenes, we investigate in a bit more detail the inner workings of the state-independent and state-dependent numerical procedures. To this end, we fix the financial scenario presented in Figure 7.3.

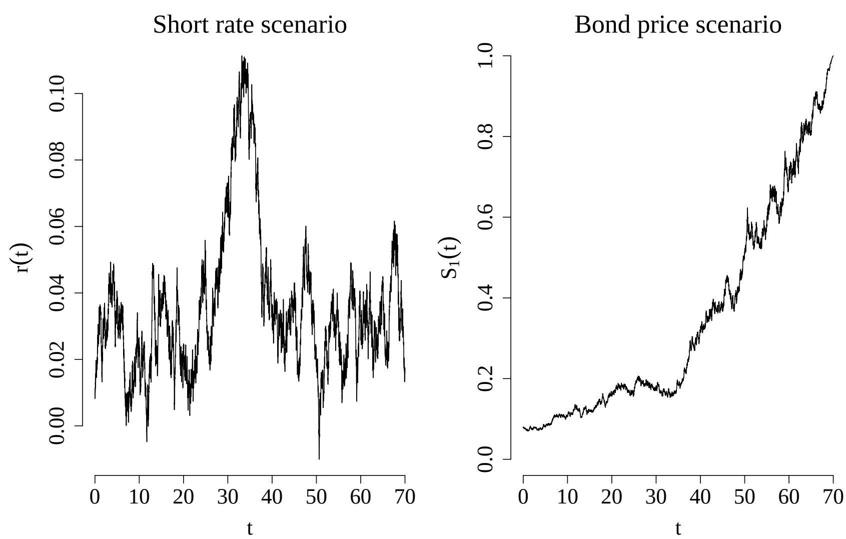

Figure 7.3: *Short rate and bond price corresponding to a single financial scenario.*

In Figure 7.4, we compare $t \mapsto p^Q_{z_0 j}(0,t)$, $j \in \mathcal{J}$, of the state-dependent implementation to $t \mapsto Q(t)p_{z_0 j}(0,t)$ of the state-independent implementation. With Proposition 7.2.2 in mind, we also compare differences in bonus cash flows across states. To be precise, Figure 7.5 contains a comparison of state-wise bonus cash flows, i.e. here we compare

$$t \mapsto p^Q_{z_0 j}(0,t)\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big)$$

of the state-dependent implementation to

$$t \mapsto Q(t)p_{z_0 j}(0,t)\Big(b^\dagger_j(t) + \sum_{\substack{k\in\mathcal{J}\\k\neq j}} b^\dagger_{jk}(t)\mu_{jk}(t)\Big)$$

of the state-independent implementation.

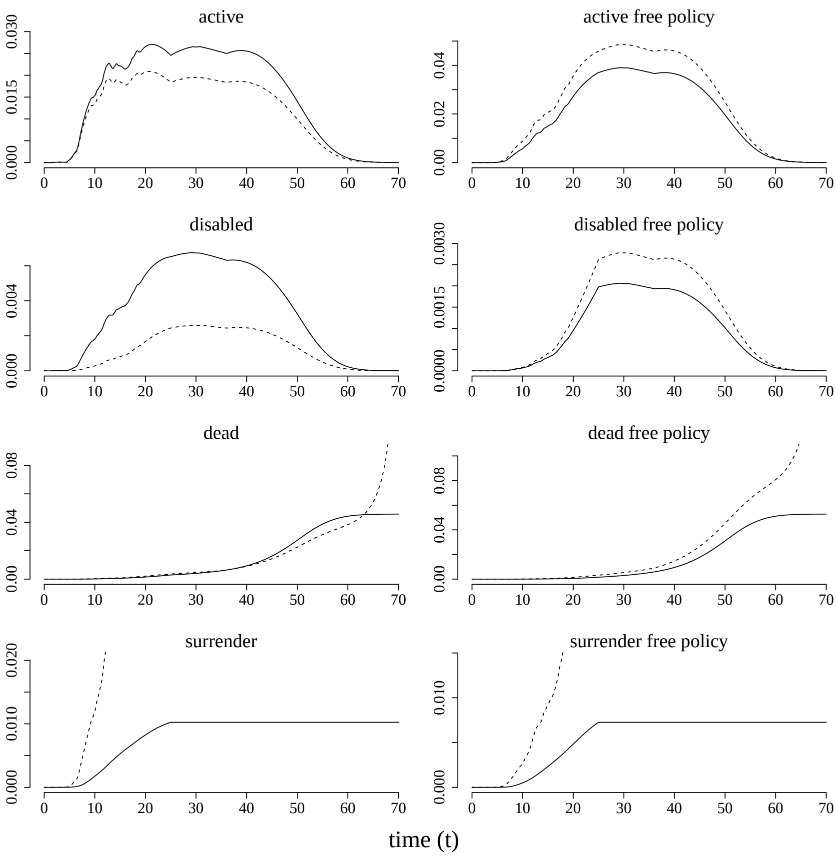

Figure 7.4: *Q-modified transition probabilities for a single financial scenario in the state-dependent implementation (solid line) and in the state-independent implementation (dashed line).*

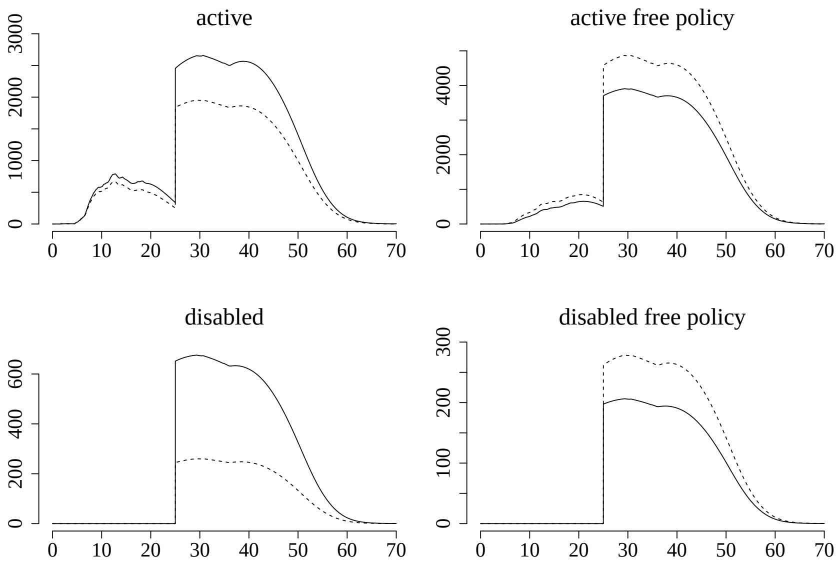

Figure 7.5: *State-wise bonus cash flows for a single financial scenario in the state-dependent implementation (solid line) and in the state-independent implementation (dashed line).*

In general, the two numerical procedures lead to fundamentally different intermediate quantities. The state-independent implementation both overestimates and underestimates – depending on the state – intermediate state-specific quantities compared to the state-dependent implementation. But intriguingly the aggregation over states cancels these differences, resulting in the same value for $V^b(0)$.

To offer a different point of view based on Remark 7.2.3 and Example 7.4.4, we take a closer look at the value of the dividend payments between implementations. To elaborate, we compare

$$t \mapsto \mathbb{E}\left[\big(r^\delta(t) - r^\star(t)\big) V^\star(t) \frac{V^\dagger_{Z(t)}(t)}{V^{\star,\dagger}_{Z(t)}(t)} \,\middle|\, \mathcal{F}^S(t) \right]$$

of the state-dependent implementation to

$$t \mapsto \mathbb{E}\left[\big(r^\delta(t) - r^\star(t)\big) \bar{V}^\star(t) \frac{V^\dagger_{Z(t)}(t)}{\bar{V}^{\star,\dagger}(t)} \,\middle|\, \mathcal{F}^S(t) \right]$$

of the state-independent implementation. The absolute relative differences are less than 1%, which indicates that it is the aggregation over states and not times that cancels the intermediate differences and gives rise to essentially identical values for $V^b(0)$.

7.6 Final remarks

In this section, we compare our methodology and results with recent advances in the literature and discuss possible extension in demand by practitioners. Subsection 7.6.1 contains comparisons with Bruhn and Lollike (2021), Falden and Nyegaard (2021), 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 7.6.3.

7.6.1 Comparison with recent advances in the literature

In Bruhn and Lollike (2021) and the follow-up paper Falden and Nyegaard (2021), 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

$$\mathbb{E}\big[\mathbb{1}_{(Z(t)=j)}V^\star(t)\,\big|\,\mathcal{F}^S(t)\big].$$

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 (2021) and Falden and Nyegaard (2021) and those of Theorem 7.3.3. 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 (7.2.17). No such assumption is made in Falden and Nyegaard (2021), which leads to more involved payment streams, although by setting $B^\dagger = B^{\circ,+}$, our payment stream equals that of Falden and Nyegaard (2021, Subsection 5.2, cf. (11)–(12)).

We consider some key concepts and provide practical insights that are not within the scope of Bruhn and Lollike (2021) and Falden and Nyegaard (2021). 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.

As discussed previously, the projection model described in Jensen and Schomacker (2015, Section 4) appears to be conceptually very close to exactly our state-independent model, cf. Subsection 7.4.2 and Example 7.4.4. 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 (2021) and Falden and Nyegaard (2021) and Jensen and Schomacker (2015).

7.6.2 State-independent approximations

The dividend strategies considered in the numerical example of Section 7.5 are to some extent absorbing in nature, meaning that they over time broadly speaking distribute the available assets to the insured. This is consistent with the fact that both the state-dependent and state-independent implementation produce a time zero market value of bonus payments identical to the available assets.

In addition to the numerical results presented in Section 7.5, we have examined other product designs and sets of parameter values. Without reporting every detail here and now, this did not immediately produce large differences in the time zero market value of bonus payments between implementations. To uncover why this is the case, further theoretical and numerical studies are required.

7.6.3 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 7.3.3 and Subsection 7.4.4, 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 7.3.3 appears to require more sophisticated methods. In the state-independent case, the assumption of affinity is not required, cf. Remark 7.4.1. 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. We are also grateful to Ann-Sophie Buchardt for her assistance with the delightful figures.

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

Bibliography

Aalen, O.O., Ø. Borgan, and H.K. Gjessing (2008). *Survival and Event History Analysis: A Process Point of View.* Statistics for Biology and Health. New York, NY: Springer New York.

Adékambi, F. and M.C. Christiansen (2017). Integral and differential equations for the moments of multistate models in health insurance. *Scandinavian Actuarial Journal* **2017**(1), pp. 29–50.

Adékambi, F. and M.C. Christiansen (2020). Probability Distributions of Multi-States Insurance Models under Semi-Markov Assumptions. *Markov Processes and Related Fields* **26**(3), pp. 517–534.

Ahmad, J. (2022). Multivariate higher order moments in multi-state life insurance. *Scandinavian Actuarial Journal* **2022**(5), pp. 399–420.

Ahmad, J. and M. Bladt (2022a). Aggregate Markov models in life insurance: estimation via the EM algorithm. *arXiv:2212.10661.*

Ahmad, J. and M. Bladt (2022b). Phase-type representations of stochastic interest rates with applications to life insurance. *arXiv:2207.11292.*

Ahmad, J., M. Bladt, and M. Bladt (2022). Estimating absorption time distributions of general Markov jump processes. *arXiv:2207.11303.*

Ahmad, J., M. Bladt, and C. Furrer (2022). Aggregate Markov models in life insurance: properties and valuation. *arXiv:2212.03705.*

Ahmad, J., K. Buchardt, and C. Furrer (2022). Computation of bonus in multi-state life insurance. *ASTIN Bulletin* **52**(1), 291–331.

Albrecher, H. and M. Bladt (2019). Inhomogeneous phase-type distributions and heavy tails. *Journal of Applied Probability* **56**(4), pp. 1044–1064.

Albrecher, H., M. Bladt, M. Bladt, and J. Yslas (2022). Mortality modeling and regression with matrix distributions. *Insurance: Mathematics and Economics* **107**, pp. 68–87.

Albrecher, H., M. Bladt, and A. Müller (2022). Joint lifetime modelling with matrix distributions. *arXiv:2207.01279*.

Albrecher, H., M. Bladt, and J. Yslas (2022). Fitting inhomogeneous phase-type distributions to data: the univariate and the multivariate case. *Scandinavian Journal of Statistics* **49**(1), pp. 44–77.

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* **18**, 19–60.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. Springer Series in Statistics. New York, NY: Springer New York.

Asmussen, S. (2003). *Applied Probability and Queues*. 2nd ed. Stochastic Modelling and Applied Probability. Springer.

Asmussen, S. and H. Albrecher (2010). *Ruin Probabilities*. 2nd ed. Advanced Series on Statistical Science & Applied Probability. World Scientific.

Asmussen, S., O. Nerman, and M. Olsson (1996). Fitting Phase-Type Distributions via the EM Algorithm. *Scandinavian Journal of Statistics* **23**(4), pp. 419–441.

Asmussen, S. and M. Steffensen (2020). *Risk and Insurance*. Series on Probability Theory and Stochastic Modelling. Springer, Cham.

Bacinello, A.R. (2001). Fair Pricing of Life Insurance Participating Policies with a Minimum Interest Rate Guaranteed. *ASTIN Bulletin* **31**(2), 275–297.

Bacinello, A.R., P. Millossovich, and A. Chen (2018). The impact of longevity and investment risk on a portfolio of life insurance liabilities. *European Actuarial Journal* **8**, pp. 257–290.

Baños, D.R. (2020). Life insurance policies with cash flows subject to random interest rate changes. *arXiv:2012.15541*.

Bauer, D., R. Kiesel, A. Kling, and J. Ruß (2006). Risk-neutral valuation of participating life insurance contracts. *Insurance: Mathematics and Economics* **39**(2), pp. 171–183.

Bharucha-Reid, A.T. (1960). *Elements of the theory of Markov processes and their applications*. McGraw-Hill series in probability and statistics. New York, NY: McGraw-Hill.

Björk, T. (2009). *Arbitrage Theory in Continuous Time*. Third edition. Oxford Finance Series. Oxford University Press.

Bladt, M., S. Asmussen, and M. Steffensen (2020). Matrix representations of life insurance payments. *European Actuarial Journal* **10**, pp. 29–67.

Bladt, M. and B.F. Nielsen (2017). *Matrix-Exponential Distributions in Applied Probability*. Probability Theory and Stochastic Modelling. Springer.

Bladt, M. and O. Peralta (2022). Strongly convergent homogeneous approximations to inhomogeneous Markov jump processes and applications. *arXiv:2204.02954*.

Bladt, M. and J. Yslas (2022). Phase-type mixture-of-experts regression for loss severities. *arXiv:2111.00581*.

Board, International Accounting Standards (2017). *IFRS 17 Insurance Contracts*. `https://www.idc.ac.il/he/specialprograms/accounting/fvf/Documents/IFRS17/WEBSITE155.pdf`.

Breuer, L. (2002). An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research* **112**(1), pp. 123–138.

Breuer, L. (2003). *From Markov jump processes to spatial queues*. Dordrecht: Springer Science & Business Media.

Brigo, D. and F. Mercurio (2006). *Interest Rate Models: Theory and Practice ; With Smile, Inflation and Credit*. 2. Aufl. Springer finance. Berlin, Heidelberg: Springer-Verlag.

Bruhn, K. and A.S. Lollike (2021). Retrospective reserves and bonus. *Scandinavian Actuarial Journal* **2021**(6), pp. 457–475.

Buchardt, K. (2014). Dependent interest and transition rates in life insurance. *Insurance: Mathematics and Economics* **55**, pp. 167–179.

Buchardt, K., C. Furrer, and T. Møller (2020). Tax- and expense-modified risk-minimization for insurance payment processes. *Scandinavian Actuarial Journal* **2020**(10), pp. 934–961.

Buchardt, K., C. Furrer, and M. Steffensen (2019). Forward transition rates. *Finance and Stochastics* **23**(4), pp. 975–999.

Buchardt, K. and T. Møller (2015). Life Insurance Cash Flows with Policyholder Behavior. *Risks* **3**(3), 290–317.

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.

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), pp. 1–25.

Christiansen, M.C. (2012). Multistate models in health insurance. *Advances in Statistical Analysis* **96**, pp. 155–186.

Christiansen, M.C. (2013). Safety margins for unsystematic biometric risk in life and health insurance. *Scandinavian Actuarial Journal* **2013**(4), pp. 286–323.

Christiansen, M.C. and B. Djehiche (2020). Nonlinear reserving and multiple contract modifications in life insurance. *Insurance: Mathematics and Economics* **93**, pp. 187–195.

Cox, S.H. and Y. Lin (2007). Natural Hedging of Life and Annuity Mortality Risks. *North American actuarial journal* **11**(3), pp. 1–15.

Diez, F. and R. Korn (2020). Yield curve shapes of Vasicek interest rate models, measure transformations and an application for the simulation of pension products. *European Actuarial Journal* **10**(1), pp. 91–120.

Elliott, R.J. and P.E. Kopp (1999). *Mathematics of financial markets*. Springer finance. Berlin: Springer.

Falden, D.K. and A.K. Nyegaard (2021). Retrospective Reserves and Bonus with Policyholder Behavior. *Risks* **9**(1).

Furrer, C. (2022). Scaled insurance cash flows: representation and computation via change of measure techniques. *Finance and Stochastics* **26**(2), pp. 359–382.

Gad, K.S.T. and J.W. Nielsen (2016). Reserves and cash flows under stochastic retirement. *Scandinavian Actuarial Journal* **2016**(10), pp. 876–904.

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.

Graham, A. (1981). *Kronecker products and matrix calculus : with applications*. Ellis Horwood series in mathematics and its applications. Chichester, NY: Ellis Horwood.

Helton, J. and S. Stuckwisch (1976). Numerical approximation of product integrals. *Journal of mathematical analysis and applications* **56**(2), pp. 410–437.

Helwich, M. (2008). *Durational effects and non-smooth semi-Markov models in life insurance*. PhD thesis. University of Rostock.

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.

Hesselager, O. and R. Norberg (1996). On probability distributions of present values in life insurance. *Insurance, mathematics and economics*. Insurance: Mathematics and Economics **18**(1), pp. 35–42.

Hoem, J.M. (1969a). Markov Chain Models in Life Insurance. *Blätter der DGVFM* **9**, pp. 91–107.

Hoem, J.M. (1969b). Purged and partial Markov chains. *Scandinavian Actuarial Journal* **1969**(3-4), pp. 147–155.

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.

Hoem, J.M. and O.O. Aalen (1978). Actuarial values of payment streams. *Scandinavian Actuarial Journal* **1978**(1), pp. 38–47.

Jacobsen, M. (2006). *Point process theory and applications: Marked point and piecewise deterministic processes*. Probability and its Applications. Birkhäuser, Boston.

Jarner, S.F. and T. Møller (2015). A partial internal model for longevity risk. *Scandinavian Actuarial Journal* **2015**(4), pp. 352–382.

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.

Jensen, N.R. (2016). Scenario-based life insurance prognoses in a multi-state Markov model. *European Actuarial Journal* **6**, pp. 307–330.

Johansen, S. (1986). Product integrals and Markov processes. *CWI Newsletter* (12), pp. 3–13.

Johnson, M.A. and M.R. Taaffe (1988). The Denseness of phase distributions. *Research Memorandum*.

Kalashnikov, V. and R. Norberg (2003). On the Sensitivity of Premiums and Reserves to Changes in Valuation Elements. *Scandinavian Actuarial Journal* **2003**(3), pp. 238–256.

Koller, M. (2012). *Stochastic Models in Life Insurance*. EAA series. Berlin, Heidelberg: Springer-Verlag.

Kurtz, T.G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of applied Probability* **7**(1), pp. 49–58.

Kurtz, T.G. (1978). Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications* **6**(3), pp. 223–240.

Latouche, G. and V. Ramaswami (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling.* ASA-SIAM Series on Statistics and Applied Mathematics. SIAM.

Levantesi, S. and M. Menzietti (2018). Natural hedging in long-term care insurance. *ASTIN Bulletin* **48**(1), 233–274.

Mijatović, A. and M. Pistorius (2013). Continuously monitored barrier options under Markov processes. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* **23**(1), pp. 1–38.

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.

Møller, T. and M. Steffensen (2007). *Market-valuation methods in life and pension insurance.* International Series on Actuarial Science. Cambridge: Cambridge University Press.

Neuts, M.F. (1978). Renewal processes of phase type. *Naval research logistics quarterly* **25**(3), pp. 445–454.

Neuts, M.F. (1981). *Matrix-geometric solutions in stochastic models: an algorithmic approach.* The Johns Hopkins Series in the Mathematical Sciences. The Johns Hopkins University Press.

Neuts, M.F. (1989). *Structured Stochastic Natrices of M/G/1 Type and Their Applications.* Marcel Dekker.

Norberg, R. (1991). Reserves in Life and Pension Insurance. *Scandinavian Actuarial Journal* **1991**, pp. 3–24.

Norberg, R. (1992). Hattendorff's theorem and Thiele's differential equation generalized. *Scandinavian Actuarial Journal* **1992**(1), pp. 2–14.

Norberg, R. (1995a). A time-continuous Markov chain interest model with applications to insurance. *Applied stochastic models and data analysis* **11**(3), pp. 245–256.

Norberg, R. (1995b). Differential equations for moments of present values in life insurance. *Insurance: Mathematics and Economics* **17**(2), pp. 171–180.

Norberg, R. (1999). A theory of bonus in life insurance. *Finance and Stochastics* **3**(4), pp. 373–390.

Norberg, R. (2001). On Bonus and Bonus Prognoses in Life Insurance. *Scandinavian Actuarial Journal* **2001**(2), pp. 126–147.

Norberg, R. (2003). The Markov Chain Market. *ASTIN Bulletin* **33**(2), 265–287.

Norberg, R (2005). Anomalous PDEs in Markov chains: Domains of validity and numerical solutions. *Finance and stochastics* **9**(4), pp. 519–537.

Norberg, R. and C.M. Møller (1996). Thiele's differential equation with stochastic interest of diffusion type. *Scandinavian Actuarial Journal* **1996**(1), pp. 37–49.

Olsson, M. (1996). Estimation of Phase-Type Distributions from Censored Data. *Scandinavian journal of statistics* **23**(4), pp. 443–460.

Ramlau-Hansen, H. (1988). Hattendorff's Theorem: A Markov chain and counting process approach. *Scandinavian Actuarial Journal* **1988**(1-3), pp. 143–156.

Ramlau-Hansen, H. (1991). Distribution of Surplus in Life Insurance. *ASTIN Bulletin* **21**, 57–71.

Rydén, T. (1994). Parameter estimation for Markov modulated Poisson processes. *Communications in Statistics: Stochastic Models* **10**(4), pp. 795–829.

Steffensen, M. (2006). Surplus-linked life insurance. *Scandinavian Actuarial Journal* **2006**, pp. 1–22.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE transactions on automatic control* **23**(3), pp. 395–404.

Wolthuis, H. (1992). Prospective and retrospective premium reserves. *Blätter der DGVFM* **20**(3), pp. 317–327.

Zaglauer, K. and D. Bauer (2008). Risk-neutral valuation of participating life insurance contracts in a stochastic interest rate environment. *Insurance: Mathematics and Economics* **43**, pp. 29–40.