Matrix representations of life insurance payments

Bladt, Mogens; Asmussen, Søren; Steffensen, Mogens

Published in:
European Actuarial Journal

DOI:
10.1007/s13385-019-00222-0

Publication date:
2020

Document version
Peer reviewed version

Citation for published version (APA):
Bladt, M., Asmussen, S., & Steffensen, M. (2020). Matrix representations of life insurance payments. *European Actuarial Journal*, *10*(1), 29-67. https://doi.org/10.1007/s13385-019-00222-0

Noname manuscript No.
(will be inserted by the editor)

Matrix representations of life insurance payments

Mogens Bladt · **Søren Asmussen** ·
Mogens Steffensen

Received: date / Accepted: date

Abstract A multi–state life insurance model is described naturally in terms of the intensity matrix of an underlying (time–inhomogeneous) Markov process which specifies the dynamics for the states of an insured person. Between and at transitions, benefits and premiums are paid, defining a payment process, and the technical reserve is defined as the present value of all future payments of the contract. Classical methods for finding the reserve and higher order moments involve the solution of certain differential equations (Thiele and Hattendorff, respectively). In this paper we present an alternative matrix–oriented approach based on general reward considerations for Markov jump processes. The matrix approach provides a general framework for effortlessly setting up general and even complex multi–state models, where moments of all orders are then expressed explicitly in terms of so–called product integrals of certain matrices. Thiele and Hattendorff type of theorems may be retrieved immediately from the matrix formulae. As a main application, methods for obtaining distributions and related properties of interest (e.g. quantiles or survival functions) of the future payments are presented from both a theoretical and practical point of view, employing Laplace transforms and methods involving orthogonal polynomials.

M. Bladt
University of Copenhagen
Department of Mathematical Sciences
Tel.: +45 35 32 07 23/24
E-mail: bladt@math.ku.dk

S. Asmussen
Aarhus University
Department of Mathematics
Tel: +45 8715 5100
E-mail: asmus@math.au.dk

M. Steffensen
University of Copenhagen
Department of Mathematical Sciences
Tel.: +45 35 32 07 23/24
E-mail: mogens@math.ku.dk

Keywords Markov reward processes; product integral; orthogonal polynomials; life-insurance; moments.

1 Introduction

In this paper we consider the distribution and moments of the total reward generated by a time inhomogeneous Markov process with a finite state space. Rewards may be earned in three different ways:

(a) accumulated according to a deterministic rate depending on the current state during sojourns.
(b) as deterministic lump sums depending on the current state, arriving according to some non–homogeneous Poisson process.
(c) at the time of transitions, deterministic lump sums depending on the states involved in the transitions, are paid out with certain (state–dependent) probabilities (which may be one or zero as well).

We are particularly interested in the case of discounted rewards which have applications in life insurance. Here the rewards (premiums and benefits) are discounted by a deterministic (though time–dependent) interest rate. The above setting is slightly more general than the standard life insurance set–up, and is inspired by finite state–space Markov additive processes (see e.g. [3], XI, 1 or [9], Chapter 10). This connection inspired us to allow for randomized lump sums as indicated under (c), and we provide in fact an example of how such a situation could arise in life insurance (see Example 2).

Our method for deriving distributions and moments uses probabilistic (sample path) arguments and matrix algebra. In particular, the matrices of interest are readily derived from the intensity matrix of the underlying Markov process and a matrix of payments. This is true for both pure and discounted rewards, where in the latter case the interest rate may be accommodated conveniently into the intensity matrix. One of our main innovations is the use of the so-called product integral. This goes back at least to [13] and has received numerous applications in statistics and Markov processes, see e.g. [20], [14]. This has been developed in general measure-theoretic contexts, but in our setting, it is basically just a formalism for the solution to systems of linear differential equations. However, the set-up is appealing by suggesting a number of properties enabling us to carry out derivations in a transparent and easy implementable way, even if it does not necessarily speed up the numerical computations.

From the ODE interpretation we retrieve both Thiele's differential equation and Hattendorff type of theorems. The moment generating function for the total (discounted) reward is obtained as a product integral involving intensity and payment matrices. All moments up to order k are then obtained by a product integral of a $(k+1)\times(k+1)$ block matrix built upon the aforementioned matrices. The product integrals may be evaluated in a number of ways. If the intensities and payments are piecewise constant, which may often be the case in practical implementations, it reduces to a product of matrix–exponentials which may be evaluated numerically by efficient methods like e.g. uniformisation, and which is available at different levels of sophistication in most software packages used in statistics or numerical mathematics.

The second moment of the total reward is of interest for example to assess whether the portfolio size is sufficiently large to justify the pricing and calculation of reserves to be based on expected values alone. Higher order moments are, however, rarely used in life insurance, but our approach provides, in principle, access to moments of all orders, and the numerical computability up to quite high orders suggest that they can also be used for approximating the cumulative distribution function (c.d.f.) of the distribution, and thereby the calculation of quantiles (values at risk or confidence intervals) which could provide valuable information concerning the actual risk. We provide a main example along this line by proposing a Gram-Charlier expansion to approximating both the density (p.d.f.) of the absolutely continuous part of the discounted future payment distribution when this is and the whole unrestricted c.d.f. The shape of the distribution can be quite challenging, particularly for the case of the p.d.f.

The idea of using multi-state (time–inhomogeneous) Markov processes as a model in life insurance dates back at least to the 1960's and was put into a modern context by [19], in which also Thiele's differential equations for the state–wise reserves are derived. A survey of this theory and some extension can be found in [6]. Variance formulas for the future payments can e.g. be found in [27] whereas for higher order moments we refer to [24]. A differential equation (Thiele) approach to calculating the c.d.f. of the discounted future payments has been considered in [18]. If one considers only unit rewards on all jumps and Poisson arrivals (no continuous rewards) in a time–homogeneous Markov jump process, then the total undiscounted reward up to time t defines a point process which is known as a Markovian Arrival Process (MAP). Apart from defining a tractable class of point process with numerous applications in applied probability, the MAPs form a simple dense class of point processes on the positive reals (see [5]).

The main contributions of the paper are as follows. The moment generating function (Laplace transform) of the total rewards (Theorem 3) generalises a similar result of [8] for time–homogeneous Markov processes. The matrix representation of moments depends on a crucial result given in Lemma 2, which generalises a similar result proved in [29] for the case of constant matrices. The reserves and moments we deal with in the analysis are the so–called partial reserves and moments, which are defined as the expected value of the (powers) of the future (discounted) payments contingent on the terminal state. These reserves and moments may well be of interest on their own. The matrix representation of all moments provides a unifying approach to the explicit solution of Thiele's and Hattendorff type of theorems. As a main application of the availability of moments up to any order, we give representations for the densities and distribution functions in terms of Gram–Chalier expansions and demonstrate the numerical feasibility of the method. In particular, this opens up for studying further distributional properties like e.g. quantiles.

The rest of the paper is organised as follows. In Section 2 we review the basic properties of the product integral, which plays an important role in the paper. The basic model and notation is set up in Section 3 and in Section 4 we use a probabilistic argument to prove a slightly extended version of Thiele's differential equation. An important technical result regarding the calculation of certain ordinary integrals via product integrals is proved in Section 5. The main construction takes place in Sections 6 and 7 where we derive explicit matrix representations for the moment generating function and higher order moments of the discounted future payments (total reward). An slightly extended version of Hattendorff's theorem is derived as a consequence in the end of Section 7. Since the moments of up to high orders are easily calculated, in Section 9 we explore the possibility of calculating the p.d.f. and c.d.f. for the total discounted future payments by means of orthogonal polynomial expansions based on central moments. In Section 9 we provide some numerical examples, and in Section 10 we conclude the paper.

2 Some relevant background

Let $\boldsymbol{A}(x) = \{a_{ij}(x)\}_{i,j=1,\dots,p}$ be a $p \times p$ matrix function. The product integral of $\boldsymbol{A}(x)$, written as

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

where $\boldsymbol{I}$ denotes the identity matrix, may be defined in a number of equivalent ways. It is e.g. the solution to Kolmogorov forward differential equation

$$\frac{\partial}{\partial s}\boldsymbol{F}(s,t) = -\boldsymbol{A}(s)\boldsymbol{F}(s,t), \quad \boldsymbol{F}(t,t) = \boldsymbol{I},$$

which by integration and repeated substitutions also yields the Peano–Baker series representation

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \boldsymbol{I} + \sum_{n=1}^{\infty} \int_s^t \int_s^{x_n} \cdots \int_s^{x_2} \boldsymbol{A}(x_1)\boldsymbol{A}(x_2)\cdots\boldsymbol{A}(x_n)\,\mathrm{d}x_1\,\mathrm{d}x_2 \cdots \mathrm{d}x_n. \quad (2.1)$$

The product integral exists if $\boldsymbol{A}(x)$ is Riemann integrable, which is henceforth assumed. From the Peano–Baker representation one may prove that

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \prod_s^x (\boldsymbol{I} + \boldsymbol{A}(u)\,\mathrm{d}u) \prod_x^t (\boldsymbol{I} + \boldsymbol{A}(u)\,\mathrm{d}u) \quad (2.2)$$

holds true for any order of s,t and x (not only $s \le x \le t$). In particular, the inverse of a product integral then exists and is given by

$$\left(\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) \right)^{-1} = \prod_t^s (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(s+t-x)\,\mathrm{d}x). \quad (2.3)$$

If the matrices $\boldsymbol{A}(x)$ all commute, then

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \exp\left(\int_s^t \boldsymbol{A}(x)\,\mathrm{d}x \right). \quad (2.4)$$

In particular, if $\boldsymbol{A}(x)=\boldsymbol{A}$ for all x, then

$$\prod_s^t (\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x) = \exp\left(\boldsymbol{A}(t-s)\right). \tag{2.5}$$

This last observation may be useful in connection with piecewise constant matrices $\boldsymbol{A}(x)$

$$\boldsymbol{A}(x)=\boldsymbol{A}_i, \quad x_{i-1}\leq x\leq x_i$$

for $i=1,2,...$ and where $x_0=0$. Then, using (2.2),

$$\prod_s^t (\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x) = \exp\left(\boldsymbol{A}_i(x_i-s)\right)\left(\prod_{k=i+1}^{j-1}\exp\left(\boldsymbol{A}_k(x_k-x_{k-1})\right)\right)\exp\left(\boldsymbol{A}_j(t-x_{j-1})\right), \tag{2.6}$$

where i and j are such that $s\in[x_{i-1},x_i]$ and $t\in[x_{j-1},x_j]$.

Matrix–exponentials can be calculated in numerous ways (see [22] and [21]) and are typically available in standard software packages at varying level of sophistication. If the exponent is an intensity (or sub–intensity matrix, i.e. row sums are non–positive) then either a Runge–Kutta method or uniformisation (see e.g. [9], p. 51) are competitive and among the most efficient.

Product integrals may also be used in the construction of time–inhomogeneous Markov processes if the primitive is an intensity matrix. Indeed, if $\boldsymbol{A}(x)$ are intensity matrices (i.e. off diagonal elements are non–negative and rows sum to 0), then their product integrals are transition matrices, and by (2.2), the Chapman–Kolmogorov' equations are then satisfied which implies the Markov property. For further details we refer to [20].

3 The general model

Consider a time–inhomogeneous Markov process $\{Z(t)\}_{t\geq 0}$ with state space $E=\{1,2,...,p\}$ and intensity matrices $\boldsymbol{M}(t)=\{\mu_{ij}(t)\}_{i,j\in E}$. Let $\boldsymbol{P}(s,t)=\{p_{ij}(s,t)\}_{i,j=1,..,p}$ denote the corresponding transition matrix. Assume that

$$\boldsymbol{M}(s)=\boldsymbol{C}(s)+\boldsymbol{D}(s),$$

where $\boldsymbol{D}(s)=\{d_{ij}(s)\}_{i,j\in E}$ denotes a $p\times p$ matrix with $d_{ij}(s)\geq 0$ and $\boldsymbol{C}(s)=\{c_{ij}(s)\}_{i,j\in E}$ is a sub–intensity matrix, i.e. its row sums are non–positive. We define a reward structure on the Markov process in the following way. At jumps from i to j, lump sums of $b^{ij}(s)$ are obtained with probability $d_{ij}(s)/(d_{ij}(s)+c_{ij}(s))$. When $Z(s)=i$ there may be two kind of rewards: a continuous rate of $b^i(s)$, so that $b^i(s)\,\mathrm{d}s$ is earned during $[s,s+\mathrm{d}s)$, and lump sums of $b^{ii}(s)$ at the events of an independent Poisson process with rate $d_{ii}(s)$ while in state i. The total reward obtained during $[s,t]$ is then given by

$$U^0(s,t) \;=\; \sum_i \bar{R}^i(s,t)+\sum_{i,j}\bar{N}^{ij}(s,t) \tag{3.1}$$

where

$$\bar{N}^{ij}(s,t) \;=\; \int_s^t b^{ij}(x)\,\mathrm{d}N^{ij}(x) \;\text{ and }\; \bar{R}^{i}(s,t) \;=\; \int_s^t b^{i}(x)1\{Z(x)=i\}\,\mathrm{d}x$$

and where $N^{ij}(x)$ for $i \neq j$ is the counting process which increases by $+1$ upon a transition of $Z(t)$ from i to j in which there is a lump sum payment, whereas $N^{ii}(x)$ is a Poisson process with rate $d_{ii}(t)$.

We assume throughout that

$$\overline{\mu} = \sup_{i,j,u} \mu_{ij}(u) < \infty\,, \quad \overline{b}^{\,\cdot} = \sup_{i,u} |b^{i}(u)| < \infty\,, \quad \overline{b}^{\,\cdot\cdot} = \sup_{i,j,u} |b^{ij}(u)| < \infty\,. \tag{3.2}$$

This ensures that all integrals and expectations in the following are well defined and finite. It does, however, also rule out the existence of a maximum age, known from e.g. de Moivre's mortality law. Although this may be considered critical from a statistical or biological point of view, it is less critical for reserving questions in actuarial science. Reserving for payments to policy holders that are close to their maximum lifetime is not really critical in a portfolio of contracts where the vast majority of policy holders are dead by then.

Our principal application is to life insurance, where the states $i \in E$ are the different conditions of an insured individual (e.g. active, unemployed, disabled or dead). Here we are interested in studying the discounted rewards, $U(s,t)$, during a time interval $[s,t]$ defined by

$$U(s,t) \;=\; \int_s^t \mathrm{e}^{-\int_s^u r}\,\mathrm{d}B(u) \;=\; \int_s^t \mathrm{e}^{-\int_s^u r(x)\,\mathrm{d}x}\,\mathrm{d}B(u)\,,$$

where $r(x) \geq 0$ is a deterministic (instantaneous) interest rate at time x and B is a payment process

$$\mathrm{d}B(t) = b^{Z(t)}(t)\,\mathrm{d}t + \sum_{j=1}^{p} b^{Z(t-)j}(t)\,\mathrm{d}N^{Z(t-)j}(t). \tag{3.3}$$

Here the continuous rates may e.g. be premiums (negative) or benefits (e.g. periodic unemployment payments). Lump sums $b^{ij}(t)$ are paid out at transitions i to j at time t with probability $d_{ij}(t)/(d_{ij}(t)+c_{ij}(t))$, while other lump sums of $b^{ii}(t)$ are paid out while the insured is in condition i at random times that will arrive according to a time–inhomogeneous Poisson process at rate $d_{ii}(t)$.

Remark 1 If $d_{ij}(t)/(d_{ij}(t)+c_{ij}(t))$ is either zero or one for all $i \neq j$, and $d_{ii}(s) = 0$ for all i, then we recover the standard multi–state Markov model in life insurance as in e.g. [19] or [23].

Remark 2 In the following we shall consider expressions of the form

$$\mathbb{E}\Big(1\{Z(t)=j\}\int_s^t \mathrm{d}B(u)\,\Big|\,Z(s)=i\Big)$$

which by (3.3) amounts to

$$\begin{aligned}
&\mathbb{E}\Big(1\{Z(t)=j\}\int_s^t \mathrm{d}B(u)\,\Big|\,Z(s)=i\Big) \\
&\quad= \sum_{k=1}^p \mathbb{E}\Big(1\{Z(t)=j\}\int_s^t 1\{Z(u)=k\}b^k(u)\,\mathrm{d}u\,\Big|\,Z(s)=i\Big) \\
&\qquad+ \sum_{k,\ell=1}^p \mathbb{E}\Big(1\{Z(t)=j\}\int_s^t 1\{Z(u)=k\}b^{k\ell}(u)\,\mathrm{d}N^{k\ell}(u)\,\Big|\,Z(s)=i\Big).
\end{aligned}$$

The $k\ell$-th element in the second sum takes the form

$$\int_s^t b^{k\ell}(u)\mathbb{E}\big(1\{Z(t)=j, Z(u)=k\}\,\mathrm{d}N^{k\ell}(u)\,\big|\,Z(s)=i\big),$$

where formally $\mathbb{E}\big(1\{Z(t)=j,Z(u)=k\}\,\mathrm{d}N^{k\ell}(u)\,\big|\,Z(s)=i\big)$ denotes the intensity measure for the process $N^{k\ell}$ conditional on $Z(s)=i$ and on the event that $Z(t)=j$. It will turn out that these (conditional) intensity measures have corresponding densities and hence can be written as $g_{k\ell}(u;i,j)\,\mathrm{d}u$ for some function g (see Lemma 1).

We are interested in calculating the moments of $U^0(s,t)$ and more generally of $U(s,t)$. To this end we define the slightly more general quantities

$$m_{ij}^{(k)} \;=\; \mathbb{E}\big[1\{Z(t)=j\}U^0(s,t)^k\,\big|\,Z(s)=i\big], \tag{3.4}$$

the corresponding matrix

$$\boldsymbol{m}^{(k)}(s,t) = \big\{m_{ij}^{(k)}\big\}_{i,j\in E}, \tag{3.5}$$

and more generally,

$$v_{ij}^{(k)}(s,t) \;=\; \mathbb{E}\big[1\{Z(t)=j\}U(s,t)^k\,\big|\,Z(s)=i\big], \tag{3.6}$$

with matrix

$$\boldsymbol{V}^{(k)}(s,t) \;=\; \big\{v_{ij}^{(k)}(s,t)\big\}_{i,j\in E} \tag{3.7}$$

for $k\in\mathbb{N}$. Define

$$\boldsymbol{B}(t) = \{b^{ij}(t)\}_{i,j\in E}, \tag{3.8}$$

$$\boldsymbol{b}(t) = (b^1(t),\dots,b^p(t))', \tag{3.9}$$

$$\boldsymbol{R}(t) = \boldsymbol{D}(t)\bullet\boldsymbol{B}(t)+\boldsymbol{\Delta}(\boldsymbol{b}(t)), \tag{3.10}$$

$$\boldsymbol{C}^{(k)}(t) = \boldsymbol{D}(t)\bullet\boldsymbol{B}^{\bullet k}(t),\ \ k\geq 2 \tag{3.11}$$

where $\boldsymbol{\Delta}(\boldsymbol{b}(t))$ denotes the diagonal matrix with the vector $\boldsymbol{b}(t)$ as diagonal, $\bullet$ is the Schur (entrywise) matrix product, i.e. $\{a_{ij}\}\bullet\{b_{ij}\}=\{a_{ij}b_{ij}\}$ and $\boldsymbol{B}^{\bullet k}(t)=\boldsymbol{B}(t)\,\bullet$

$\cdots \bullet \boldsymbol{B}(t)$ (k terms). Thus $\boldsymbol{B}(t)$ is the matrix which arranges all lump sums, $\boldsymbol{\Delta}(\boldsymbol{b}(t))$ is the diagonal matrix containing the rates of benefit in the different states, and

$$\boldsymbol{R}(t) = \begin{pmatrix} d_{11}(t)b^{11}(t)+b^1(t) & d_{12}(t)b^{12}(t) & \cdots & d_{1p}(t)b^{1p}(t) \\ d_{21}(t)b^{21}(t) & d_{22}(t)b^{22}(t)+b^2(t) & \cdots & d_{2p}(t)b^{2p}(t) \\ \vdots & \vdots & \vdots & \vdots \\ d_{p1}(t)b^{p1}(t) & d_{p2}(t)b^{p2}(t) & \cdots & d_{pp}(t)b^{pp}(t)+b^p(t) \end{pmatrix}$$

accommodates the total rates of earnings in the different states and transitions. The matrices $\boldsymbol{C}^{(k)}(t)$ are used for higher order moments.

The state-wise prospective reserves are defined as $\mathbb{E}[U(s,t)|Z(s)=i]$ for all $i \in E$, which are then the elements of the vector $\boldsymbol{V}^{(1)}(s,t)\boldsymbol{e}$, where $\boldsymbol{e}$ is the column vector of ones (see (3.7)). We say that the matrix $\boldsymbol{V}^{(1)}(s,t)$ contains the *partial (statewise prospective) reserves* and we refer to the matrix itself as such. Though the partial reserve may have its own merit, it is introduced primarily for mathematical convenience.

4 Partial reserves and Thiele's differential equations

First we start with an integral representation of the first order moment $\boldsymbol{m}^{(1)}(s,t)$. In the proof and in the rest of the paper we use, as is common practice in applied probability, an infinitesimal formalism for carrying out probabilistic arguments. Readers uncomfortable with this approach may replace du by h and adding a term $o(h)$, where $o(h)$ is a function for which $o(h)/h \to 0$ as $h \to 0$.

Lemma 1 $\boldsymbol{m}^{(1)}(s,t) \;=\; \int_s^t \boldsymbol{P}(s,u)\boldsymbol{R}(u)\boldsymbol{P}(u,t)\,\mathrm{d}u\,.$

Proof The (i,j)-th element of $\boldsymbol{m}^{(1)}(s,t)$ can be decomposed in to sums of continuous linear rewards, jump rewards and rewards from Poisson arrivals (see also Remark 2), i.e.

$$\begin{aligned} \mathbb{E}\big(1\{Z(t)=j\}U^0(s,t)\,\big|\,Z(s)=i\big) \;&=\; \mathbb{E}\Big(1\{Z(t)=j\}\int_s^t \mathrm{d}B(u)\,\Big|\,Z(s)=i\Big) \\ &= \sum_k \int_s^t b^k(u)\mathbb{E}\big(1\{Z(t)=j\}1\{Z(u)=k\}\,\big|\,Z(s)=i\big)\,\mathrm{d}u \\ &\quad + \sum_{k,\ell}\int_s^t b^{k\ell}(u)\mathbb{E}\big(1\{Z(t)=j\}\,\mathrm{d}N^{k\ell}(u)\,\big|\,Z(s)=i\big) \end{aligned}$$

Consider a pure jump reward, $k \neq \ell$. Then by the Markov property,

$$\begin{aligned} &\mathbb{E}\big(1\{Z(t)=j\}\,\mathrm{d}N^{k\ell}(u)\,\big|\,Z(s)=i\big) \\ &\quad=\; p_{ik}(s,u)\frac{d_{k\ell}(u)}{\mu_{k\ell}(u)}\mu_{k\ell}(u)\,\mathrm{d}u\,p_{\ell j}(u,t) \;=\; \boldsymbol{e}_i'\boldsymbol{P}(s,u)\boldsymbol{e}_k d_{k\ell}(u)\,\mathrm{d}u\,\boldsymbol{e}_\ell'\boldsymbol{P}(u,t)\boldsymbol{e}_j. \end{aligned}$$

since the probability of a transition from k to ℓ in $[u,u+\mathrm{d}u)$ is $\mu_{k\ell}(u)\,\mathrm{d}u$, which is accompanied by a reward of $b^{k\ell}(u)$ with probability

$$\frac{d_{k\ell}(u)}{d_{k\ell}(u)+c_{k\ell}(u)}=\frac{d_{k\ell}(u)}{\mu_{k\ell}(u)}\,.$$

If $\mu_{k\ell}(u)=0$, the transition will be redundant and the resulting probability is hence zero (i.e. $0/0=0$). Hence the expected reward in $[s,t]$ due to jumps from k to ℓ amount to

$$\int_s^t \mathbb{E}\left(b^{k\ell}(u)1\{Z(t)=j\}\,\mathrm{d}N^{k\ell}(u)|Z(s)=i\right)$$
$$=\int_s^t \boldsymbol{e}_i'\boldsymbol{P}(s,u)\boldsymbol{e}_k d_{k\ell}(u)b^{k\ell}(u)\boldsymbol{e}_\ell'\boldsymbol{P}(u,t)\boldsymbol{e}_j\,\mathrm{d}u.$$

This settles the case for the jumps. Concerning the linear rewards, consider state k. Here its contribution amounts to

$$\int_s^t b^k(u)\mathbb{P}\big(Z(t)=j,Z(u)=k\,|\,Z(s)=i\big)\,\mathrm{d}s=\int_s^t b^k(u)\boldsymbol{e}_i'\boldsymbol{P}(s,u)\boldsymbol{e}_k\boldsymbol{e}_k'\boldsymbol{P}(u,t)\boldsymbol{e}_j\,\mathrm{d}s.$$

For the case of lump sums from Poisson arrivals in state k, the contribution is

$$\int_s^t \boldsymbol{e}_i'\boldsymbol{P}(s,u)\boldsymbol{e}_k d_{kk}(u)b^{k\ell}(u)\boldsymbol{e}_k'\boldsymbol{P}(u,t)\boldsymbol{e}_j\,\mathrm{d}u\,.$$

The total reward is then obtained by summing over the three different types of reward, which in matrix notation exactly yields the result. □

Next we consider the partial reserve. Following standard notation from life insurance, we define the partial reserve for a fixed time horizon (the terminal date) T (which can be $+\infty$) as a function of t only as

$$\boldsymbol{V}(t)=\boldsymbol{V}^{(1)}(t,T). \tag{4.1}$$

We denote the elements of $\boldsymbol{V}(t)$ by $v_{ij}(t)$.

In addition to (3.2), we now assume that $r(u)$ and $\boldsymbol{M}(u)$ are piecewise continuous with at most a finite number of discontinuity points. The product integrals of theses will then be differentiable with derivatives which are continuous in all but a possibly finite number of points, and the interchange of differentiation and integration in the following is then warranted by Leibniz' integration rule (see e.g. [26], p. 447). Then we have the following Thiele type of differential equation for the partial reserve.

Theorem 1 (Thiele) $\boldsymbol{V}(t)=\boldsymbol{V}^{(1)}(t,T)$ *satisfies*

$$\frac{\partial}{\partial t}\boldsymbol{V}(t)=r(t)\boldsymbol{V}(t)-\boldsymbol{M}(t)\boldsymbol{V}(t)-\boldsymbol{R}(t)\boldsymbol{P}(t,T)$$

with terminal condition $\boldsymbol{V}(T)=\boldsymbol{0}$.

Proof From Lemma 1 we have that

$$\mathbb{E}\big[\mathrm{d}B(u)1\{Z(T)=j\}\,\big|\,Z(t)=i\big] \;=\; \boldsymbol{e}_i'\boldsymbol{P}(t,u)\boldsymbol{R}(u)\boldsymbol{P}(u,T)\boldsymbol{e}_j\,\mathrm{d}u.$$

Hence we get that

$$\begin{aligned} v_{ij}(t) &= \int_t^T \mathrm{e}^{-\int_s^u r(u)\,\mathrm{d}u}\mathbb{E}\big[1\{Z(T)=j\}\,\mathrm{d}B(u)\,\big|\,Z(t)=i\big] \\ &= \int_t^T \mathrm{e}^{-\int_t^u r(s)\,\mathrm{d}s}\boldsymbol{e}_i'\boldsymbol{P}(t,u)\boldsymbol{R}(u)\boldsymbol{P}(u,T)\boldsymbol{e}_j\,\mathrm{d}u \\ &= \boldsymbol{e}_i'\int_t^T \prod_t^u (1-r(s)\,\mathrm{d}s)\prod_t^u(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u\,\boldsymbol{e}_j \\ &= \boldsymbol{e}_i'\int_t^T \prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u\,\boldsymbol{e}_j. \end{aligned}$$

In matrix notation,

$$\boldsymbol{V}(t) = \int_t^T \prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u. \tag{4.2}$$

Thus

$$\begin{aligned} \frac{\partial}{\partial t}\boldsymbol{V}(t) &= \frac{\partial}{\partial t}\int_t^T \prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u \\ &= \int_t^T \left[\frac{\partial}{\partial t}\prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\right]\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u \\ &\qquad -\boldsymbol{I}\cdot\boldsymbol{R}(t)\prod_t^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s)) \\ &= -(\boldsymbol{M}(t)-r(t)\boldsymbol{I}))\int_t^T \prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u \\ &\qquad -\boldsymbol{R}(t)\prod_t^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s)) \\ &= -(\boldsymbol{M}(t)-r(t)\boldsymbol{I}))\boldsymbol{V}(t)-\boldsymbol{R}(t)\prod_t^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s)) \\ &= -(\boldsymbol{M}(t)-r(t)\boldsymbol{I}))\boldsymbol{V}(t)-\boldsymbol{R}(t)\boldsymbol{P}(t,T), \end{aligned}$$

i.e.

$$\frac{\partial}{\partial t}\boldsymbol{V}(t) = r(t)\boldsymbol{V}(t)-\boldsymbol{M}(t)\boldsymbol{V}(t)-\boldsymbol{R}(t)\boldsymbol{P}(t,T), \tag{4.3}$$

with obvious boundary condition $\boldsymbol{V}(T)=\boldsymbol{0}$. □

As an immediate consequence, using that $\boldsymbol{P}(t,T)\boldsymbol{e}=\boldsymbol{e}$ being a transition probability matrix, we recover the usual Thiele differential equation.

Corollary 1 *The vector of prospective reserves $\boldsymbol{V}_{Th}(t) = \boldsymbol{V}^{(1)}(t,T)\boldsymbol{e}$ satisfies*

$$\frac{\partial}{\partial t}\boldsymbol{V}_{Th}(t) = r(t)\boldsymbol{V}_{Th}(t) - \boldsymbol{M}(t)\boldsymbol{V}_{Th}(t) - \boldsymbol{R}(t)\boldsymbol{e},$$

with terminal condition $\boldsymbol{V}_{Th}(T) = \boldsymbol{0}$.

5 Matrix representation of the reserve

In this section we will provide a matrix representation of the reserve. We start with an important general result which extends a result by [29] from matrix–exponentials to product integrals. Here the matrices $\boldsymbol{A}(x)$ and $\boldsymbol{C}(x)$ must be square but not necessarily of the same dimension.

Lemma 2 *Let $\boldsymbol{A}(x)$, $\boldsymbol{B}(x)$ and $\boldsymbol{C}(x)$ be continuous matrix functions. Then we have that*

$$\prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{A}(x) & \boldsymbol{B}(x) \\ \boldsymbol{0} & \boldsymbol{C}(x) \end{pmatrix} \mathrm{d}x\right) = \begin{pmatrix} \prod_s^t(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x) & \int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u \\ \boldsymbol{0} & \prod_s^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x) \end{pmatrix}.$$

Proof By Leibniz' integration rule,

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u \\
&= \int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\frac{\partial}{\partial t}\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u + \prod_s^t(\boldsymbol{I}+\boldsymbol{A}(u))\boldsymbol{B}(t)\prod_t^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x) \\
&= \int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u\,\boldsymbol{C}(t) + \prod_s^t(\boldsymbol{I}+\boldsymbol{A}(u))\boldsymbol{B}(t).
\end{aligned}$$

Let $\boldsymbol{B}(s,t)$ denote the matrix on the right hand side in the Lemma. Then

$$\begin{aligned}
\frac{\partial}{\partial t}\boldsymbol{B}(s,t) &= \begin{pmatrix} \frac{\partial}{\partial t}\prod_s^t(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x) & \frac{\partial}{\partial t}\int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u \\ \boldsymbol{0} & \frac{\partial}{\partial t}\prod_s^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x) \end{pmatrix} \\
&= \begin{pmatrix} \prod_s^t(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x) & \int_s^t \prod_s^u(\boldsymbol{I}+\boldsymbol{A}(x)\,\mathrm{d}x)\boldsymbol{B}(u)\prod_u^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x)\,\mathrm{d}u \\ \boldsymbol{0} & \prod_s^t(\boldsymbol{I}+\boldsymbol{C}(x)\,\mathrm{d}x) \end{pmatrix} \begin{pmatrix} \boldsymbol{A}(t) & \boldsymbol{B}(t) \\ \boldsymbol{0} & \boldsymbol{C}(t) \end{pmatrix} \\
&= \boldsymbol{B}(s,t)\begin{pmatrix} \boldsymbol{A}(t) & \boldsymbol{B}(t) \\ \boldsymbol{0} & \boldsymbol{C}(t) \end{pmatrix}.
\end{aligned}$$

Also, $\boldsymbol{B}(t,t) = \boldsymbol{I}$. Therefore,

$$\boldsymbol{B}(s,t) = \prod_s^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{A}(u) & \boldsymbol{B}(u) \\ \boldsymbol{0} & \boldsymbol{C}(u) \end{pmatrix} \mathrm{d}u\right).$$

□

Remark 3 It is enough to assume that $\boldsymbol{B}(x)$ and $\boldsymbol{C}(x)$ are continuous to carry out the proof above. Also, if we carry out the proof of Lemma 2 differentiating w.r.t. s instead of t, we see that we may assume continuity of $\boldsymbol{A}(x)$ and $\boldsymbol{B}(x)$ instead.

Remark 4 Lemma 2 also holds for functions with a finite number of discontinuities, which can be seen be decomposing the product integrals into product over intervals where the function is continuous and using (2.2).

Consider an $np \times np$ block matrix on the form

$$\boldsymbol{A}(x) = \begin{pmatrix} \boldsymbol{A}_{11}(x) & \boldsymbol{A}_{12}(x) & \cdots & \boldsymbol{A}_{1n}(x) \\ \boldsymbol{0} & \boldsymbol{A}_{22}(x) & \cdots & \boldsymbol{A}_{2n}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{3n}(x) \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{nn}(x) \end{pmatrix}$$

and write

$$\boldsymbol{B}(x) = \prod_s^t (\boldsymbol{I} + \boldsymbol{A}(x)\,\mathrm{d}x) = \begin{pmatrix} \boldsymbol{B}_{11}(s,t) & \boldsymbol{B}_{12}(s,t) & \cdots & \boldsymbol{B}_{1n}(s,t) \\ \boldsymbol{0} & \boldsymbol{B}_{22}(s,t) & \cdots & \boldsymbol{B}_{2n}(s,t) \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{B}_{3n}(s,t) \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{B}_{nn}(s,t) \end{pmatrix},$$

which we know from Lemma 2 must be an upper triangular block matrix. Then using Lemma 2 again, we get

$$\big(\boldsymbol{B}_{12}(s,t), \boldsymbol{B}_{13}(s,t), ..., \boldsymbol{B}_{1n}(s,t)\big) =$$
$$\int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{11}(u)\,\mathrm{d}u)\,[\boldsymbol{A}_{12}(x), ..., \boldsymbol{A}_{1n}(x)] \prod_x^t \left(\boldsymbol{I} + \begin{pmatrix} \boldsymbol{A}_{22}(u) & \boldsymbol{A}_{23}(u) & \cdots & \boldsymbol{A}_{2n}(u) \\ \boldsymbol{0} & \boldsymbol{A}_{33}(u) & \cdots & \boldsymbol{A}_{3n}(u) \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{4n}(u) \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{nn}(u) \end{pmatrix} \mathrm{d}u \right) \mathrm{d}x$$

so that

$$\boldsymbol{B}_{1n}(s,t) = \int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{11}(u)\,\mathrm{d}u)\,[\boldsymbol{A}_{12}(x), ..., \boldsymbol{A}_{1n}(x)] \begin{pmatrix} \boldsymbol{B}_{2n}(x,t) \\ \boldsymbol{B}_{3n}(x,t) \\ \vdots \\ \boldsymbol{B}_{nn}(x,t) \end{pmatrix} \mathrm{d}x$$

which can be written as

$$\boldsymbol{B}_{1n}(s,t) = \sum_{i=2}^n \int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{11}(u)\,\mathrm{d}u) \boldsymbol{A}_{1i}(x) \boldsymbol{B}_{in}(x,t)\,\mathrm{d}x.$$

Applying Lemma 2 to $\boldsymbol{A}(x)$ from downwards and up in the block partitioning, we then get that

$$\begin{aligned}\boldsymbol{B}_{nn}(s,t) &= \prod_s^t(\boldsymbol{I}+\boldsymbol{A}_{nn}(x)\,\mathrm{d}x)\\ \boldsymbol{B}_{n-1,n}(s,t) &= \int_s^t \prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{n-1,n-1}(u)\,\mathrm{d}u)\boldsymbol{A}_{n-1,n}(x)\prod_x^t(\boldsymbol{I}+\boldsymbol{A}_{nn}(u)\,\mathrm{d}u)\\ &= \int_s^t \prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{n-1,n-1}(u)\,\mathrm{d}u)\boldsymbol{A}_{n-1,n}(x)\boldsymbol{B}_{nn}(x,t)\,\mathrm{d}x\end{aligned}$$

and more generally,

$$\boldsymbol{B}_{i,n}(s,t) = \sum_{j=i+1}^{n}\int_s^t \prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{ii}(u)\,\mathrm{d}u)\boldsymbol{A}_{ij}(x)\boldsymbol{B}_{jn}(x,t)\,\mathrm{d}x. \tag{5.1}$$

Theorem 2 *The partial reserve defined in* (4.1)*,* $\boldsymbol{V}(t)$*, (and the transition matrix) can be calculated through the product integral*

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

Proof Applying Lemma 2 to (4.2) we get that

$$\begin{aligned}&\prod_t^T\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{M}(u)-r(u)\boldsymbol{I} & \boldsymbol{R}(u)\\ \boldsymbol{0} & \boldsymbol{M}(u)\end{pmatrix}\mathrm{d}u\right) =\\ &\begin{pmatrix}\prod_t^T(\boldsymbol{I}+(\boldsymbol{M}(u)-r(u)\boldsymbol{I})\,\mathrm{d}u) & \int_t^T\prod_t^u(\boldsymbol{I}+(\boldsymbol{M}(s)-r(s)\boldsymbol{I})\,\mathrm{d}s)\boldsymbol{R}(u)\prod_u^T(\boldsymbol{I}+\boldsymbol{M}(s)\,\mathrm{d}s))\,\mathrm{d}u\\ \boldsymbol{0} & \prod_t^T(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u).\end{pmatrix}\\ &= \begin{pmatrix}\prod_t^T(\boldsymbol{I}+(\boldsymbol{M}(u)-r(u)\boldsymbol{I})\,\mathrm{d}u) & \boldsymbol{V}(t)\\ \boldsymbol{0} & \boldsymbol{P}(t,T).\end{pmatrix}.\end{aligned} \tag{5.2}$$

□

So both the partial reserve $\boldsymbol{V}(t)$ and $\boldsymbol{P}(t,T)=\{p_{ij}(t,T)\}$ are calculated trough one single calculation of the product integral. This is convenient if we are interested in the calculation of the expected future payments conditional on $Z(t)=i$ and $Z(T)=j$, since

$$\mathbb{E}\big[U(t,T)\,\big|\,Z(t)=i,Z(T)=j\big] = \frac{v_{ij}(t)}{p_{ij}(t,T)} = \frac{v_{ij}^{(1)}(t,T)}{p_{ij}(t,T)}. \tag{5.3}$$

Such quantities are side results that may be of interest on their own, see the conclusion concerning ideas of application. The reserve notion formalized in (5.3) bears a resemblance with a particular notion of a retrospective reserve introduced in [23]. There, various versions of retrospective reserves are studied, including one that evaluates the past payments, at a given time point, conditional on the initial and the current states of Z, respectively. This is similar to the conditioning occurring in (5.3). A key difference is that this is studied as a function of the terminal time point (in our case time T) exclusively ina resemblance with a particular notion of a retrospective reserve introduced in [23], leading to conditioning on a non-monotonic series of sigma-algebras, whereas we study the numerator isolatedly and as a function of the initial time point (in our case t) exclusively. This lead to fundamental differences in structures of results. Further, only the first moment of such a retrospective reserve is considered ina resemblance with a particular notion of a retrospective reserve introduced in [23].

6 Moment generating function of rewards and future payments

Recall the definition (3.1) of the total reward $U^0(s,t)$ which is the undiscounted future payments in an insurance context. Let

$$F_{ij}(x;s,t) = \mathbb{P}(Z(t)=j, U^0(s,t) \le x \mid Z(s)=i),$$
$$F_{ij}^*(\theta;s,t) = \int_{-\infty}^{\infty} \mathrm{e}^{\theta x}\,\mathrm{d}F_{ij}(x;s,t) = \mathbb{E}\left[\mathrm{e}^{\theta U^0(s,t)} 1\{Z(t)=j\} \,\middle|\, Z(s)=i\right],$$
$$\boldsymbol{F}^*(\theta;s,t) = \{F_{ij}^*(\theta;s,t)\}_{i,j=1,\dots,p}.$$

This is well defined and finite for all real θ due to the assumption of (3.2). In fact, then $|U^0(s,t)| \le (T-s)\overline{b}^{\,\cdot} + \overline{b}^{\,\cdot\cdot} N$ where N is Poisson$(\overline{\mu}(T-s))$. Obviously, assumption (3.2) is extremely weak in the insurance context.

Theorem 3 *The distribution of the total reward $U^0(s,t)$ has a moment generating function given by*

$$\boldsymbol{F}^*(\theta;s,t) = \prod_s^t \left(\boldsymbol{I} + \left[\boldsymbol{D}(u) \bullet \{\mathrm{e}^{\theta b^{k\ell}(u)}\}_{k,\ell} + \boldsymbol{C}(u) + \theta \boldsymbol{\Delta}(\boldsymbol{b}(u))\right] \mathrm{d}u\right). \tag{6.1}$$

Proof For $s \le u \le t$, $U^0(s,t) = U^0(s,u) + U^0(u,t)$, and hence

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{e}^{\theta U^0(s,t)} 1\{Z(t)=j\} \,\middle|\, Z(s)=i\right] \\
&= \sum_k \mathbb{E}\left[\mathrm{e}^{\theta U^0(s,u)} \mathrm{e}^{\theta U^0(u,t)} 1\{Z(u)=k\} 1\{Z(t)=j\} \,\middle|\, Z(s)=i\right] \\
&= \sum_k \mathbb{E}\left(\mathrm{e}^{\theta U^0(s,u)} 1\{Z(u)=k\} \,\middle|\, Z(s)=i\right) \mathbb{E}\left(\mathrm{e}^{\theta U^0(u,t)} 1\{Z(t)=j\} \,\middle|\, Z(u)=k\right),
\end{aligned}$$

where the last equality follows by conditioning on $Z(u)$ and using the Markov property. In matrix notation this amounts to

$$\boldsymbol{F}^*(\theta;s,t) = \boldsymbol{F}^*(\theta;s,u)\boldsymbol{F}^*(\theta;u,t).$$

Consider $\boldsymbol{F}^*(\theta;t,t+\mathrm{d}t)$. Its ijth element equals

$$\begin{aligned} F^*_{ij}(\theta;t,t+\mathrm{d}t) &= (\delta_{ij}+c_{ij}(t)\,\mathrm{d}t)\mathrm{e}^{\theta b^i(t)\,\mathrm{d}t}+d_{ij}(t)\,\mathrm{d}t\mathrm{e}^{\theta b^{ij}(t)} \\ &= (\delta_{ij}+c_{ij}(t)\,\mathrm{d}t)(1+\theta b^i(t)\,\mathrm{d}t)+d_{ij}(t)dt\mathrm{e}^{\theta b^{ij}(t)} \\ &= \delta_{ij}+c_{ij}(t)\,\mathrm{d}t+\theta\delta_{ij}b^i(t)\,\mathrm{d}t+d_{ij}(t)\,\mathrm{d}t\mathrm{e}^{\theta b^{ij}(t)} \end{aligned}$$

so that

$$\boldsymbol{F}^*(\theta;t,t+\mathrm{d}t)=\boldsymbol{I}+\boldsymbol{C}(t)\,\mathrm{d}t+\theta\boldsymbol{\Delta}(\boldsymbol{b}(t))\,\mathrm{d}t+\boldsymbol{D}(t)\bullet\left\{\mathrm{e}^{\theta b^{k\ell}(t)}\right\}_{k,\ell}\mathrm{d}t.$$

Hence

$$\begin{aligned} \boldsymbol{F}^*(\theta;s,t+\mathrm{d}t)-\boldsymbol{F}^*(\theta;s,t) &= \boldsymbol{F}^*(\theta;s,t)\,(\boldsymbol{F}^*(\theta;t,t+\mathrm{d}t)-\boldsymbol{I}) \\ &= \boldsymbol{F}^*(\theta;s,t)\left[\boldsymbol{C}(t)\,\mathrm{d}t+\theta\boldsymbol{\Delta}(\boldsymbol{b}(t))\,\mathrm{d}t+\boldsymbol{D}(t)\bullet\left\{\mathrm{e}^{\theta b^{k\ell}(t)}\right\}_{k,\ell}\mathrm{d}t\right], \end{aligned}$$

i.e.

$$\frac{\partial}{\partial t}\boldsymbol{F}^*(\theta;s,t)=\boldsymbol{F}^*(\theta;s,t)\left[\boldsymbol{C}(t)+\theta\boldsymbol{\Delta}(\boldsymbol{b}(t))+\boldsymbol{D}(t)\bullet\left\{\mathrm{e}^{\theta b^{k\ell}(t)}\right\}_{k,\ell}\right] \tag{6.2}$$

and together with the obvious boundary condition $\boldsymbol{F}^*(\theta;s,t)=\boldsymbol{I}$, the result then follows. □

7 Higher order moments

Define the matrices

$$\boldsymbol{F}^{(k)}(x)=\begin{pmatrix} \boldsymbol{M}(x) & \binom{k}{1}\boldsymbol{R}(x) & \binom{k}{2}\boldsymbol{C}^{(2)}(x) & \cdots & \binom{k}{k-1}\boldsymbol{C}^{(k-1)}(x) & \boldsymbol{C}^{(k)}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x)\boldsymbol{I} & \binom{k-1}{1}\boldsymbol{R}(x) & \cdots & \binom{k-1}{k-2}\boldsymbol{C}^{(k-2)}(x) & \boldsymbol{C}^{(k-1)}(x) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{M}(x) & \boldsymbol{R}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix} \tag{7.1}$$

and

$$\boldsymbol{H}^{(k)}(s,t)=\begin{pmatrix} \boldsymbol{P}(s,t) & \binom{k}{1}\boldsymbol{m}^{(1)}(s,t) & \binom{k}{2}\boldsymbol{m}^{(2)}(s,t) & \cdots & \binom{k}{k-1}\boldsymbol{m}^{(k-1)}(s,t) & \boldsymbol{m}^{(k)}(s,t) \\ \boldsymbol{0} & \boldsymbol{P}(s,t) & \binom{k-1}{1}\boldsymbol{m}^{(1)}(s,t) & \cdots & \binom{k-1}{k-2}\boldsymbol{m}^{(k-2)}(s,t) & \boldsymbol{m}^{(k-1)}(s,t) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{P}(s,t) & \boldsymbol{m}^{(1)}(s,t) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{P}(s,t) \end{pmatrix} \tag{7.2}$$

Then we have the following main result.

Theorem 4

$$\prod_s^t (\boldsymbol{I}+\boldsymbol{F}^{(k)}(x)\,\mathrm{d}x) = \boldsymbol{H}^{(k)}(s,t).$$

Proof First we notice that $\boldsymbol{F}^*(0;s,t) = \prod_s^t(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u) = \boldsymbol{P}(s,t)$, and we can obtain (recall (3.5))

$$\boldsymbol{m}^{(k)}(s,t) \;=\; \Big\{\mathbb{E}\Big[1\{Z(t)=j\}U^0(s,t)^k \,\Big|\, Z(s)=i\Big]\Big\}_{i,j}$$

by

$$\boldsymbol{m}^{(k)}(s,t) = \frac{\partial^k}{\partial\theta^k}\boldsymbol{F}^*(\theta;s,t)\bigg|_{\theta=0}.$$

Now

$$\frac{\partial^k}{\partial\theta^k}\{\mathrm{e}^{\theta b^{ij}(t)}\}_{i,j}\bigg|_{\theta=0} = \boldsymbol{B}(t)^{\bullet k} = \boldsymbol{B}(t)\bullet\cdots\bullet\boldsymbol{B}(t)$$

(k factors) whereas for $k=0$ (no differentiation, only evaluation at $\theta=0$) it equals the matrix which has all entrances equal to one. Differentiating (6.1) first with respect to s and then with respect to θ we get that

$$\begin{aligned}\frac{\partial^k}{\partial\theta^k}\frac{\partial}{\partial s}\boldsymbol{F}^*(\theta;s,t) &= \frac{\partial^k}{\partial\theta^k}\Big(-\Big[\boldsymbol{D}(s)\bullet\{\mathrm{e}^{\theta b^{ij}(s)}\}_{i,j}+\boldsymbol{C}(s)+\theta\boldsymbol{\Delta}(\boldsymbol{b}(s))\}\Big]\boldsymbol{F}^*(\theta;s,t)\Big)\\ &= -\sum_{m=0}^{k}\binom{k}{m}\frac{\partial^m}{\partial\theta^m}\Big[\boldsymbol{D}(s)\bullet\{\mathrm{e}^{\theta b^{ij}(s)}\}_{i,j}+\boldsymbol{C}(s)+\theta\boldsymbol{\Delta}(\boldsymbol{b}(s))\}\Big]\frac{\partial^{k-m}}{\partial\theta^{k-m}}\boldsymbol{F}^*(\theta;s,t).\end{aligned}$$

Recalling that

$$\boldsymbol{R}(s) = \boldsymbol{D}(s)\bullet\boldsymbol{B}(s)+\boldsymbol{\Delta}(\boldsymbol{b}(s))$$

and since

$$\Big[\boldsymbol{D}(s)\bullet\{\mathrm{e}^{\theta b^{ij}(s)}\}_{i,j}+\boldsymbol{C}(s)+\theta\boldsymbol{\Delta}(\boldsymbol{b}(s))\}\Big]\bigg|_{\theta=0} = \boldsymbol{D}(s)+\boldsymbol{C}(s) = \boldsymbol{M}(s)$$

we get that

$$\begin{aligned}\frac{\partial}{\partial s}\boldsymbol{m}^{(k)}(s,t) \;=\; &-\Big[\boldsymbol{M}(s)\boldsymbol{m}^{(k)}(s,t)+k\boldsymbol{R}(s)\boldsymbol{m}^{(k-1)}(s,t)\\ &+\sum_{m=2}^{k}\binom{k}{m}\boldsymbol{D}(s)\bullet\boldsymbol{B}^{\bullet m}(s)\boldsymbol{m}^{(k-m)}(s,t)\Big],\end{aligned} \tag{7.3}$$

where

$$\boldsymbol{m}^{(0)}(s,t) = \boldsymbol{F}^*(0;s,t) = \prod_s^t(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u) = \boldsymbol{P}(s,t).$$

Multiplying from the left on both sides with $\prod_t^s(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u)$ (see also (2.3)) we get

$$\begin{aligned}\frac{\partial}{\partial s}\left(\prod_t^s(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u)\boldsymbol{m}^{(k)}(s,t)\right) \;=\; &-k\prod_t^s(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u)\boldsymbol{R}(s)\boldsymbol{m}^{(k-1)}(s,t)\\ &-\sum_{m=2}^{k}\binom{k}{m}\prod_t^s(\boldsymbol{I}+\boldsymbol{M}(u)\,\mathrm{d}u)\,(\boldsymbol{D}(s)\bullet\boldsymbol{B}^{\bullet m}(s))\,\boldsymbol{m}^{(k-m)}(s,t).\end{aligned}$$

Integrating this equation then gives

$$\begin{aligned}\boldsymbol{m}^{(k)}(s,t) &= k\int_s^t \boldsymbol{P}(s,x)\boldsymbol{R}(x)\boldsymbol{m}^{(k-1)}(x,t)\,\mathrm{d}x \\ &\quad + \sum_{m=2}^{k}\binom{k}{m}\int_s^t \boldsymbol{P}(s,x)\left(\boldsymbol{D}(x)\bullet\boldsymbol{B}^{\bullet m}(x)\right)\boldsymbol{m}^{(k-m)}(x,t)\,\mathrm{d}x \\ &\overset{(3.11)}{=} k\int_s^t \boldsymbol{P}(s,x)\boldsymbol{R}(x)\boldsymbol{m}^{(k-1)}(x,t)\,\mathrm{d}x \\ &\quad + \sum_{m=2}^{k}\binom{k}{m}\int_s^t \boldsymbol{P}(s,x)\boldsymbol{C}^{(m)}(x)\boldsymbol{m}^{(k-m)}(x,t)\,\mathrm{d}x. \end{aligned} \tag{7.4}$$

Now we employ an induction argument to prove the general identity. For $k=1$ the result amounts to

$$\prod_s^t\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{M}(u) & \boldsymbol{R}(u)\\ \boldsymbol{0} & \boldsymbol{M}(u)\end{pmatrix}\mathrm{d}u\right)=\begin{pmatrix}\boldsymbol{P}(s,t) & \boldsymbol{m}^{(1)}(s,t)\\ \boldsymbol{0} & \boldsymbol{P}(s,t)\end{pmatrix},$$

which indeed holds true since Lemma 2 implies that

$$\boldsymbol{m}^{(1)}(s,t)=\int_s^t \boldsymbol{P}(s,x)\boldsymbol{R}(x)\boldsymbol{P}(x,t)\,\mathrm{d}x,$$

which has been previously established in Lemma 1. Assume that the results hold true for dimension $k-1$. Partition the matrix $\boldsymbol{F}^{(k)}(u)$ as

$$\boldsymbol{F}^{(k)}(u)=\begin{pmatrix}\boldsymbol{M}(u) & \boldsymbol{x}^{(k)}(u)\\ \boldsymbol{0} & \boldsymbol{F}^{(k-1)}(u)\end{pmatrix},$$

where

$$\boldsymbol{x}^{(k)}(u)=\left(\binom{k}{1}\boldsymbol{R},\ \binom{k}{2}\boldsymbol{C}^{(2)},\ \binom{k}{3}\boldsymbol{C}^{(3)},\cdots,\ \binom{k}{k-1}\boldsymbol{C}^{(k-1)},\ \binom{k}{k}\boldsymbol{C}^{(k)}\right).$$

Then use Lemma 2, the induction hypothesis and (7.4) to verify the correct form of the first block row. □

Remark 5 The central moments

$$\mathbb{E}\left[\left(U(s,t)-\mathbb{E}\left[U(s,t)\,|Z(s)=i\right]\right)^k\,\middle|\,Z(s)=i\right]$$

can be obtained by Theorem 4 by a simple re-parametrisation, replacing $\boldsymbol{b}(t)=(b^1(t),\dots,b^p(t))$ by

$$\boldsymbol{b}(t)-\frac{\mathbb{E}\left[U(s,t)\,\middle|\,Z(s)=i\right]}{t-s}\boldsymbol{e},$$

where the expected values $\mathbb{E}\left[U(s,t)\,\middle|\,Z(s)=i\right]$ are then first calculated by Theorem 4 in the usual way with $k=1$. □

Next we turn to the case of discounted rewards (future payments). In principle we may calculate the moments of the future discounted payments by applying Theorem 4 to the discounted rewards

$$\mathrm{e}^{-\int_s^u r} b^i(u) = \mathrm{e}^{-\int_s^u r(x)\,\mathrm{d}x} b^i(u) \quad \text{and} \quad \mathrm{e}^{-\int_s^u r} b^{ij}(u) = \mathrm{e}^{-\int_s^u r(x)\,\mathrm{d}x} b^{ij}(u).$$

We may however obtain a more explicit matrix representation which involves the interest rate $r(x)$ in a more convenient way, and which is closer to standard intuition in life insurance (like e.g. Hattendorff's theorem). We define $\boldsymbol{F}_U^{(k)}(x)$ as the matrix

$$\begin{pmatrix} \boldsymbol{M}(x)-kr(x)\boldsymbol{I} & \binom{k}{1}\boldsymbol{R}(x) & \binom{k}{2}\boldsymbol{C}^{(2)}(x) & \cdots & \binom{k}{k-1}\boldsymbol{C}^{(k-1)}(x) & \boldsymbol{C}^{(k)}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x)-(k-1)r(x)\boldsymbol{I} & \binom{k-1}{1}\boldsymbol{R}(x) & \cdots & \binom{k-1}{k-2}\boldsymbol{C}^{(k-2)}(x) & \boldsymbol{C}^{(k-1)}(x) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{M}(x)-r(x)\boldsymbol{I} & \boldsymbol{R}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix} \tag{7.5}$$

and let

$$\boldsymbol{G}^{(k)}(s,t) = \prod_s^t (\boldsymbol{I} + \boldsymbol{F}_U^{(k)}(x)\,\mathrm{d}x).$$

The matrix $\boldsymbol{F}_U^{(k)}(x)$ is a $(k+1)p \times (k+1)p$ block–partitioned matrix with blocks of sizes $p \times p$. Thus $\boldsymbol{G}^{(k)}(s,t)$ is also a $(k+1)p \times (k+1)p$ matrix, and we define a similar block partitioning as for $\boldsymbol{F}^{(k)}(x)$, letting $\boldsymbol{G}_{ij}^{(k)}(s,t)$ denote the ij'th block which corresponds to the ij'th block of $\boldsymbol{F}_U^{(k)}(x)$. Then we have the following main result.

Theorem 5 *For $j = 1,\dots,k$ we have that*

$$\boldsymbol{V}^{(j)}(s,t) = \boldsymbol{G}_{k+1-j,k+1}^{(k)}(s,t),$$

whereas

$$\boldsymbol{P}(s,t) = \boldsymbol{G}_{k+1,k+1}.$$

The theorem states that the right block–column of $\boldsymbol{G}^{(k)}(s,t)$ contains the moments $\boldsymbol{V}^{(j)}(s,t)$, starting with the highest moment in the upper right corner and finishing with the transition matrix in the lower right corner (which may be considered as the zeroth moment). Symbolically,

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{F}^{(k)}(x)\,\mathrm{d}x) = \begin{pmatrix} * * * * \cdots * \; \boldsymbol{V}^{(k)}(s,t) \\ * * * * \cdots * \; \boldsymbol{V}^{(k-1)}(s,t) \\ * * * * \cdots * \; \boldsymbol{V}^{(k-2)}(s,t) \\ \vdots \; \vdots \; \vdots \; \vdots \; \vdots \quad \vdots \qquad \vdots \\ * * * * \cdots * \; \boldsymbol{V}^{(1)}(s,t) \\ * * * * \cdots * \; \boldsymbol{P}(s,t) \end{pmatrix}. \tag{7.6}$$

The idea of the general proof is most easily explained through the following example, which proves the result of Theorem 5 for the case $k=2$, which is the lowest non–trivial order.

Example 1 (Quadratic moment) First we consider the product integral

$$\boldsymbol{G}(s,t)=\prod_s^t\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{A}_{11}(x) & \boldsymbol{A}_{12}(x) & \boldsymbol{A}_{13}(x)\\ \boldsymbol{0} & \boldsymbol{A}_{22}(x) & \boldsymbol{A}_{23}(x)\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{33}(x)\end{pmatrix}\mathrm{d}x\right)=\begin{pmatrix}\boldsymbol{G}_{11}(s,t) & \boldsymbol{G}_{12}(s,t) & \boldsymbol{G}_{13}(s,t)\\ \boldsymbol{0} & \boldsymbol{G}_{22}(s,t) & \boldsymbol{G}_{23}(s,t)\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{G}_{33}(s,t)\end{pmatrix}.$$

Employing Lemma 2 inductively, by first partioning the matrix as

$$\left(\begin{array}{c|cc}\boldsymbol{A}_{11}(x) & \boldsymbol{A}_{12}(x) & \boldsymbol{A}_{13}(x)\\ \hline \boldsymbol{0} & \boldsymbol{A}_{22}(x) & \boldsymbol{A}_{23}(x)\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{33}(x)\end{array}\right),$$

we see that

$$\boldsymbol{G}_{11}(s,t)=\prod_s^t(\boldsymbol{I}+\boldsymbol{A}_{11}(x)\,\mathrm{d}x)$$

and

$$\begin{aligned}\begin{pmatrix}\boldsymbol{G}_{22}(s,t) & \boldsymbol{G}_{23}(s,t)\\ \boldsymbol{0} & \boldsymbol{G}_{33}(s,t)\end{pmatrix}&=\prod_s^t\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{A}_{22}(x) & \boldsymbol{A}_{23}(x)\\ \boldsymbol{0} & \boldsymbol{A}_{33}(x)\end{pmatrix}\mathrm{d}x\right)\\ &=\begin{pmatrix}\prod_s^t(\boldsymbol{I}+\boldsymbol{A}_{22}(x)\,\mathrm{d}x) & \int_s^t\prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{22}(u)\,\mathrm{d}u)\boldsymbol{A}_{23}(x)\prod_x^t(\boldsymbol{I}+\boldsymbol{A}_{11}(u)\,\mathrm{d}u)\,\mathrm{d}x\\ \boldsymbol{0} & \prod_s^t(\boldsymbol{I}+\boldsymbol{A}_{33}(x)\,\mathrm{d}x)\end{pmatrix}\end{aligned}$$

whereas

$$\begin{aligned}&(\boldsymbol{G}_{12}(s,t),\boldsymbol{G}_{13}(s,t))\\ &\quad=\int_s^t\prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{11}(u)\,\mathrm{d}u)\,(\boldsymbol{A}_{12}(x),\ \boldsymbol{A}_{13}(x))\prod_x^t\left(\boldsymbol{I}+\begin{pmatrix}\boldsymbol{A}_{22}(u) & \boldsymbol{A}_{23}(u)\\ \boldsymbol{0} & \boldsymbol{A}_{33}(u)\end{pmatrix}\mathrm{d}u\right)\mathrm{d}x\end{aligned}$$

so that

$$\boldsymbol{G}_{12}(s)=\int_s^t\prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{11}(u)\,\mathrm{d}u)\boldsymbol{A}_{12}(x)\prod_x^t(\boldsymbol{I}+\boldsymbol{A}_{22}(u)\,\mathrm{d}u)\,\mathrm{d}x$$

and

$$\begin{aligned}\boldsymbol{G}_{13}(s)&=\int_s^t\prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{11}(u)\,\mathrm{d}u)\boldsymbol{A}_{13}(x)\prod_x^t(\boldsymbol{I}+\boldsymbol{A}_{33}(u)\,\mathrm{d}u)\,\mathrm{d}x\\ &\quad+\int_s^t\prod_s^x(\boldsymbol{I}+\boldsymbol{A}_{11}(u)\,\mathrm{d}u)\boldsymbol{A}_{12}(x)\int_x^t\prod_x^y(\boldsymbol{I}+\boldsymbol{A}_{22}(u)\,\mathrm{d}u)\boldsymbol{A}_{23}(y)\prod_y^t(\boldsymbol{I}+\boldsymbol{A}_{33}(u)\,\mathrm{d}u)dy.\end{aligned}$$

Now assume that we are concerned with the discounted prices. Then at any time $x\in[s,t]$, we scale the prices by $\mathrm{e}^{-\int_s^x r}$. In the above expression for $\boldsymbol{G}_{13}(s,t)$,

$$\boldsymbol{A}_{13}(x)=\boldsymbol{D}(x)\bullet\tilde{\boldsymbol{B}}(x)\bullet\tilde{\boldsymbol{B}}(x)$$

while

$$\boldsymbol{A}_{12}(x) = \boldsymbol{A}_{23}(x) = \boldsymbol{D}(x) \bullet \tilde{\boldsymbol{B}}(x) + \boldsymbol{\Delta}(\tilde{\boldsymbol{b}}(x)),$$

where

$$\tilde{\boldsymbol{B}}(x) = \mathrm{e}^{-\int_s^x r}\boldsymbol{B}(x) \text{ and } \tilde{\boldsymbol{b}}(x) = \mathrm{e}^{-\int_s^x r}\boldsymbol{b}(x).$$

In the expression for $\boldsymbol{G}_{13}(s,t)$, there will be a scaling factor of $\boldsymbol{A}_{13}(x)$ of size $\mathrm{e}^{-2\int_s^x r}$, ditto in $\boldsymbol{A}_{12}(x)$ amounts to $\mathrm{e}^{-\int_s^x r}$, while the factor in $\boldsymbol{A}_{23}(y)$ is given by $\mathrm{e}^{-\int_s^y r} = \mathrm{e}^{-\int_s^x r}\mathrm{e}^{-\int_x^y r}$. Thus we may write

$$\begin{aligned} \boldsymbol{G}_{13}(s,t) = & \int_s^t \prod_s^x (\boldsymbol{I} + [\boldsymbol{A}_{11}(u) - 2r(u)\boldsymbol{I}]\,\mathrm{d}u)\boldsymbol{A}_{13}(x) \prod_x^t (\boldsymbol{I} + \boldsymbol{A}_{33}(u)\,\mathrm{d}u)\,\mathrm{d}x \\ & + \int_s^t \prod_s^x (\boldsymbol{I} + [\boldsymbol{A}_{11}(u) - 2r(u)\boldsymbol{I}]\,\mathrm{d}u)\boldsymbol{A}_{12}(x) \\ & \cdot \int_x^t \prod_x^y (\boldsymbol{I} + [\boldsymbol{A}_{22}(u) - r(u)\boldsymbol{I}]\,\mathrm{d}u)\boldsymbol{A}_{23}(y) \prod_y^t (\boldsymbol{I} + \boldsymbol{A}_{33}(u)\,\mathrm{d}u)\,\mathrm{d}y\,\mathrm{d}x. \end{aligned}$$

Let

$$\boldsymbol{H}^{(2)}(x) = \begin{pmatrix} \boldsymbol{M}(x) - 2r(x)\boldsymbol{I} & 2\boldsymbol{R}(x) & \boldsymbol{C}^{(2)}(x) \\ \boldsymbol{0} & \boldsymbol{M}(x) - r(x)\boldsymbol{I} & \boldsymbol{R}(x) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M}(x) \end{pmatrix},$$

and

$$\boldsymbol{V}^{(2)}(s,t) = \prod_s^t (\boldsymbol{I} + \boldsymbol{H}^{(2)}(x)\,\mathrm{d}x) = \begin{pmatrix} \boldsymbol{V}_{11}^{(2)}(s,t) & \boldsymbol{V}_{12}^{(2)}(s,t) & \boldsymbol{V}_{13}^{(2)}(s,t) \\ \boldsymbol{0} & \boldsymbol{V}_{22}^{(2)}(s,t) & \boldsymbol{V}_{23}^{(2)}(s,t) \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{V}_{33}^{(2)}(s,t) \end{pmatrix}.$$

Then

$$\begin{aligned} \boldsymbol{V}_{33}^{(2)}(s,t) &= \boldsymbol{P}(s,t) \\ \boldsymbol{V}_{23}^{(2)}(s,t) &= \{\mathbb{E}\left(1\{Z(t)=j\}\boldsymbol{U}(s,t)|Z(s)=i\right)\}_{i,j} \\ \boldsymbol{V}_{13}^{(2)}(s,t) &= \left\{\mathbb{E}\left(1\{Z(t)=j\}\boldsymbol{U}^2(s,t)\middle|Z(s)=i\right)\right\}_{i,j}. \end{aligned}$$

In particular,

$$\begin{aligned} \mathbb{E}\left[U(s,t)|Z(s)=i\right] &= \boldsymbol{e}_i'\boldsymbol{V}_{23}^{(2)}(s,t)\boldsymbol{e} \\ \mathbb{E}\left[U^2(s,t)|Z(s)=i\right] &= \boldsymbol{e}_i'\boldsymbol{V}_{13}^{(2)}(s,t)\boldsymbol{e} \end{aligned}$$

□

We now turn to the general proof.

Proof (of Theorem 5) We apply Theorem 4 to the discounted prices $e^{-\int_s^u r} b^i(u)$ and $e^{-\int_s^u r} b^{ij}(u)$. This will indeed provide us with the correct result (for fixed s), and as in Example 1 we redistribute the discounted terms into the block diagonal matrices. For simplicity of identification of the individual blocks of the matrix, we write $\boldsymbol{F}_U^{(k)}(u)$ in a block partitioned way as

$$\boldsymbol{F}_U^{(k)}(u) = \begin{pmatrix} \boldsymbol{A}_{11}(u) & \boldsymbol{A}_{12}(u) & \cdots & \boldsymbol{A}_{1,k+1}(u) \\ \boldsymbol{0} & \boldsymbol{A}_{22}(u) & \cdots & \boldsymbol{A}_{2,k+1}(u) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_{k+1,k+1}(u) \end{pmatrix}$$

and

$$\prod_s^t (\boldsymbol{I} + \boldsymbol{F}_U^{(k)}(u)\,\mathrm{d}u) = \begin{pmatrix} \boldsymbol{B}_{11}(s,t) & \boldsymbol{B}_{12}(s,t) & \cdots & \boldsymbol{B}_{1,k+1}(s,t) \\ \boldsymbol{0} & \boldsymbol{B}_{22}(s,t) & \cdots & \boldsymbol{B}_{2,k+1}(s,t) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{B}_{k+1,k+1}(s,t) \end{pmatrix}.$$

For example, $\boldsymbol{A}_{ii}(u) = \boldsymbol{M}(u)$ and $\boldsymbol{A}_{i,i+1}(u) = (k-i+1)\boldsymbol{R}(u)$. The matrix $\boldsymbol{A}_{i,i+m}(x)$ is then scaled by $e^{-m\int_s^x r}$. For $k = 1$ it is clear that

$$\boldsymbol{B}_{k,k+1}(s,t) = \int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{kk}(u)\,\mathrm{d}u)\boldsymbol{A}_{k,k+1}(x) \prod_x^t (\boldsymbol{I} + \boldsymbol{A}_{k+1,k+1}(u)\,\mathrm{d}s)\ \mathrm{d}x$$

has scaling factor $e^{-\int_s^x r}$, while in Example 1 we saw that $\boldsymbol{B}_{k-1,k+1}(s,t)$ has scaling factor $e^{-2\int_s^x r}$. Now assume (induction) that all $\boldsymbol{B}_{j,k+1}(s,t)$, $j = i+1,...,k$ have scaling factors $\exp(-(k-j+1)\int_s^x r)$. From the recursion (5.1),

$$\boldsymbol{B}_{i,k+1}(s,t) = \sum_{j=i+1}^{k+1} \int_s^t \prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{ii}(u)\,\mathrm{d}u)\boldsymbol{A}_{ij}(x)\boldsymbol{B}_{j,k+1}(x,t)\,\mathrm{d}x,$$

we see that $\boldsymbol{A}_{ij}(x)$ produces a scaling factor $\exp(-(j-i)\int_s^x r()$, while $\boldsymbol{B}_{j,k+1}(x,t)$ can be written as another integral over x to t with integration variable y, say, which will then have scaling factors (induction) of size $\exp(-(k-j+1)\int_s^y r)$. Now write

$$\exp\left(-(k-j+1)\int_s^y r\right) = \exp\left(-(k-j+1)\int_s^x r\right)\exp\left(-(k-j+1)\int_x^y r\right)$$

and pull out the factor $\exp(-(k-j+1)\int_s^x r()$ to get a scaling factor of

$$\exp\left(-(j-i)\int_s^x r\right) = \exp\left(-(k-j+1)\int_s^x r\right)\exp\left(-(k-i+1)\int_s^x r(\right).$$

This scaling factor can then be put together with $\prod_s^x (\boldsymbol{I} + \boldsymbol{A}_{ii}(u)\,\mathrm{d}u)$, which in turn is the (i,i) block level entrance which is simply $\prod_s^x (\boldsymbol{I} + \boldsymbol{M}(u)\,\mathrm{d}u)$. □

We may then obtain a slightly generalized version of Hattendorff's theorem.

Theorem 6

$$\frac{\partial}{\partial s}\boldsymbol{V}^{(k)}(s,t)$$
$$= (kr(s)\boldsymbol{I}-\boldsymbol{M}(s))\boldsymbol{V}^{(k)}(s,t)-k\boldsymbol{R}(s)\boldsymbol{V}^{(k-1)}(s,t)-\sum_{i=2}^{k}\binom{k}{i}\boldsymbol{C}^{(i)}(s)\boldsymbol{V}^{(k-i)}(s,t),$$

with terminal condition $\boldsymbol{V}^{(k)}(t,t)=\mathbf{0}$.

Proof Follows from differentiation of $\prod_s^t(\boldsymbol{I}+\boldsymbol{F}^{(k)}(x)\,\mathrm{d}x)$ with respect to s, obtaining a Kolmogorov type of differential equation, with $\boldsymbol{F}^{(k)}(x)$ given by (7.1), and comparing to (7.6). We only need the first block row of $\boldsymbol{F}^{(k)}(x)$ and the last block column of $\prod_s^t(\boldsymbol{I}+\boldsymbol{F}^{(k)}(x)\,\mathrm{d}x)$. □

This theorem reduces to the state-wise standard Hattendorff theorem for kth order moments, which is achieved by post–multiplying the differential equation by the vector $\boldsymbol{e}=(1,1,...,1)'$.

8 Gram-Charlier expansions of the full distribution

The c.d.f. or density of $X=U(s,T)$ can of course be evaluated by Laplace transform inversion from Theorem 3, say implementing via the Euler or Post–Widder methods (see [1]). However, the procedure is numerically challenging and somewhat tedious since the Laplace transforms may take on both very large and very small values for different arguments. Given the availability of all moments, an attractive alternative is Gram-Charlier expansions via orthogonal polynomials.

The method can briefly be summarized as follows. Consider a reference density $f_0(x)$ having all moments $\int x^k f_0(x)\,\mathrm{d}x$ well-defined and finite, and a target density $f(x)$ for which all moments $\mathbb{E}[X^k]=\int x^k f(x)\,\mathrm{d}x$ can be computed. Consider $L_2(f_0)$ with inner product $\langle g,h\rangle=\int g(x)h(x)f_0(x)\,\mathrm{d}x$ and let $p_0(x),p_1(x),\ldots$ be a set of orthonormal polynomials, i.e. $\langle p_n,p_m\rangle=\delta_{nm}$. If this set is complete in $L_2(f_0)$ and

$$f/f_0\in L_2(f_0)\,,\quad \text{i.e.}\ \int\frac{f^2(x)}{f_0(x)}\,\mathrm{d}x<\infty\quad \text{or equivalently}\ f^2/f_0\in L_1(\mathrm{Leb}),\tag{8.1}$$

we can then expand f/f_0 in the p_n to get

$$f(x)\;=\;f_0(x)\Big\{1+\sum_{n=1}^{\infty}c_np_n(x)\Big\}\quad\text{where}\ c_n\;=\;\langle f/f_0,p_n\rangle\;=\;\mathbb{E}[p_n(X)]\,.\tag{8.2}$$

If the emphasis is on the c.d.f. F or the quantiles, simply integrate this to get an expansion of $F(x)$.

For fast convergence of the series (8.2), f_0 should be chosen as much alike f as possible. The most popular choice is the normal density with the same mean μ and variance σ^2 as f, in which case $c_1=c_2=0$ (one has always $c_0=1$). This implies

$p_n(x) = H_n\big((x-\mu)/\sigma\big)/\sqrt{n!}$ for $n \geq 1$ where H_n is the nth (probabilistic) Hermite polynomial defined by $(\mathrm{d}^n/\mathrm{d}x^n)\mathrm{e}^{-x^2/2} = (-1)^n H_n(x)\mathrm{e}^{-x^2/2}$. In particular, with

$$d_n = \frac{1}{n!}\int_{-\infty}^{\infty} H_n\big((x-\mu)/\sigma\big) f(x)\,\mathrm{d}x$$

and

$$f_0(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{(x-\mu)^2/2\sigma^2},$$

we have

$$f(x) \;=\; f_0(x)\Big\{1 + \sum_{n=3}^{\infty} d_n H_n\big((x-\mu)/\sigma\big)\Big\}, \tag{8.3}$$

$$F(x) \;=\; \Phi\big((x-\mu)/\sigma\big) - \sigma f_0(x)\sum_{n=3}^{\infty} d_n H_{n-1}\big((x-\mu)/\sigma\big). \tag{8.4}$$

The conditions for (8.4) to be a valid expansion are in fact just

$$f/f_0^{1/2} \in L_1(\mathrm{Leb}), \quad \text{i.e.} \int \mathrm{e}^{(x-\mu)^2/4\sigma^2} f(x)\,\mathrm{d}x < \infty, \tag{8.5}$$

cf. [12, p. 223]. Truncated versions of (8.2) or (8.4) go under the name of Edgeworth expansions; the examples with the whole series not converging simply arise when conditions (8.1) or (8.5) is violated (whereas completeness holds when f_0 is normal). See, e.g., [28], [12, p. 133, 222ff.] and [7] for more detail. Actuarial applications of the method and related expansions can be found, e.g., in [10], [2], [16], [17], [4], [15].

When implementing the method in the insurance context $X = U(s,T)$, a difficulty is that absolute continuity typically fails. More precisely, the target distribution F is a mixture of certain atoms $a_1, a_2, \ldots$ with probability masses $q_1, q_2, \ldots.$, and an absolute continuous part thus having a density. One then has to take $f(x)$ as the (conditional) density of the absolutely continuous part,

$$f(x)\mathrm{d}x \;=\; \mathbb{P}\big(X \in \mathrm{d}x \,\big|\, X \neq a_1, X \neq a_2, \ldots\big).$$

Most often, there is only one atom with q_1 easily computable. Examples of atoms:

1) the initial state i is held throughout $(s,T]$, occuring w.p. $q_i = \int_s^T \mathrm{e}^{\int_s^T \mu_{ii}}$, so that $a_1 = \int_s^T \mathrm{e}^{-r(u-s)} b^i(u)\,\mathrm{d}u$.

2) No discounting and equal lump sum payments, $b^{ij}(t) \equiv b^{ij}$, $r = 0$. Then $U(s,T)$ is a linear combination of the b^{ij}.

These are more or less the only natural ways to get atoms that occur to us (but see Remark 8 below). For simplicity of notation, we assume there is only one atom and write $a = a_1$, $q = q_1$ (the modifications in the case of several atoms are trivial).

To implement the Gram-Charlier expansion of f, define

$$m_1 = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, \quad m_j = \int_{-\infty}^{\infty} (x-m_1)^j f(x)\,\mathrm{d}x,\; j = 2,3,\ldots$$

Obviously,

$$\mathbb{E}\big[(U(s,T)-m_1)^j\big] \;=\; q(a-m_1)^j+(1-q)m_j, \quad j\geq 2 \tag{8.6}$$

whereas for $j=1$,

$$\mathbb{E}\big[U(s,T)-m_1\big] \;=\; q(a-m_1). \tag{8.7}$$

The program for computing an α-quantile z_α is then the following:

1. Compute $\mathbb{E}U(s,T)$ via Theorem 7.1 with $k=1$ and compute
$$m_1 \;=\; \frac{\mathbb{E}U(s,T)-qa}{1-q}\,.$$
2. Choose $k>1$ and compute $\mathbb{E}\big[(U(s,T)-m_1)^j\big]$ for $j=2,\ldots,k$ via Theorem 7.1. To this end, replace the drift parameter $b^i(t)$ by
$$b^i(t):=\begin{cases} b^i(t)-\dfrac{m_1}{T-s} & \text{if } r=0\\[2ex] b^i(t)-\dfrac{r\,m_1}{1-\exp(-r*(T-s))} & \text{if } r\neq 0. \end{cases}$$
Solve next (8.6) for the m_j to get
$$m_j=\frac{\mathbb{E}\big[(U(s,T)-m_1)^j\big]-q(a-m_1)^j}{1-q}.$$
3. Take f_0 as the normal density with mean m_1 and variance $\sigma_f^2=m_2-m_1^2$. Write $H_n(x)=\sum_0^n a_{j;n}x^j$ and compute the
$$d_n \;=\; \frac{1}{n!}\sum_{j=0}^{n}\frac{a_{j;n}m_j}{\sigma_f^j}\,, \quad n=3,\ldots,k$$
4. Approximate the conditional density f and the unconditional c.d.f. F by
$$\widehat{f}_k(x) \;=\; f_0(x)\Big\{1+\sum_{n=3}^{k}d_nH_n\big((x-m_1)/\sigma_f\big)\Big\}\,,$$
$$\widehat{F}_k(x) \;=\; q1_{x\geq a}+(1-q)\left[\Phi\big((x-m_1)/\sigma_f\big)-\sigma_f f_0(x)\sum_{n=3}^{k}d_nH_{n-1}\big((x-m_1)/\sigma_f\big)\right].$$
5. Solve $\widehat{F}_k(z_\alpha^k)=\alpha$ to get a candidate z_α^k for z_α.
6. Repeat from step 2) with a larger k until z_α^k stabilizes.

At the formal mathematical level, one needs to verify the L_2 condition (8.1). We call a model with constant intensities $\mu_{ij}(t)\equiv\mu_{ij}$ *feed-forward* if there are no loops, i.e. no chain $i_0i_1\ldots i_N$ with $i_0=i_N$ and all $\mu_{i_{n-1}i_n}(u_n)>0$ for some $u_1<\cdots<u_N$. In such a case, with $T<\infty$, $U(s,T)$ has finite support, so because a normal f_0 is bounded below on compact intervals it would suffice that f is bounded. These condition are not uncommonly satisfied in life insurance though we may also have cases with $T=+\infty$ when pensions are involved (but see Remark 6 below). However, (8.1) may actually fail in complete generality. This and further aspects of condition (8.1) is discussed in the Appendix. The following result seems, however, sufficient for all practical purposes.

Theorem 7 *Assume that all intensities and rewards are piecewise constant, that $T < \infty$, that the distribution of $U(s,T]$ has an absolutely continuous part with conditional density f and that f_0 is normal(μ, σ_f^2). Assume furthermore that $d_{ii}(s) = 0$ for all i, i.e. there are no lump sums due to Poisson arrivals. Then the L_2 condition* (8.1) *holds for f if either* (i) *the model is feed-forward,* (ii) $b_k^{ij} = 0$ *for all $i \neq j$ or, more generally,* (iii) $|b_k^{ij}| > 0$ *implies that there is no path from j to i, i.e. no chain $i_0 i_n \dots i_N$ with $i_0 = j, i_N = i$ and all $\mu_{i_{n-1}i_n} > 0$.*

Proof Let $\widetilde{f}(x)$ be the unconditional density and

$$\begin{aligned} F(A\,|\,s,t;i,j) &= \mathbb{P}\Big(\int_s^t \mathrm{e}^{-\int_s^u r(u)\,\mathrm{d}u}\mathrm{d}B(u) \in A, Z(t) = j \,\Big|\, Z(s) = i\Big) \\ &= F^{\mathrm{ac}}(A\,|\,s,t;i,j) + F^{\mathrm{at}}(A\,|\,s,t;i,j), \end{aligned}$$

where $F^{\mathrm{ac}}, F^{\mathrm{at}}$ are the absolutely continuous, resp. atomic, parts (due to the special structure, there can be no singular but non-atomic part). Let $g(x\,|\,s,t;i,j)$ be the density of F^{ac} so that $\widetilde{f}(x) = \sum_j g(x\,|\,0,T;i,j)$ when $Z(0) = i$. Let $\overline{g}(s,t;i,j) = \sup_x g(x\,|\,s,t;i,j)$ and assume it shown that $\overline{g}(0,T-1;i,j) < \infty$, $\overline{g}(T-1,T;i,j) < \infty$ for all i,j. We then get

$$\begin{aligned} g(0,T;i,j) &= \sum_k \int g(x-y\,|\,0,T-1;i,k)F(\mathrm{d}y\,|\,T-1,T;k,j) \\ &\qquad + \sum_k \int g(x-y\,|\,T-1,T;k,j)F(\mathrm{d}y\,|\,0,T-1;i,k) \\ &\leq \sum_k \overline{g}(0,T-1;i,k) \int F(\mathrm{d}y\,|\,T-1,T;k,j) \\ &\qquad + \sum_k \overline{g}(T-1,T;k,j) \int F(\mathrm{d}y\,|\,0,T-1;i,k) \end{aligned}$$

so that $\overline{g}(0,T;i,j) < \infty$. Thus we may assume that $T = 1$ and simply write $\mu_{ij,k} = \mu_{ij}$ etc.

The stated conditions imply in all three cases that f has finite support. Indeed, the support is contained in $[-A,A]$ where $A = p\max|b^i| + p(p-1)\max|b^{ij}|$ in cases (i) or (iii) and $A = p\max|b^i|$ in case (ii). Since f_0 is bounded below on $[-A,A]$ for any A, it thus suffices to show that f is bounded.

Assume first that all $b^{ij} = 0$, , and let x be fixed. Define $S_N \subset [0,1]^N$ as $S_N = \{0 < t_1 < \cdots < t_N < 1\}$ and $h_n = t_n - t_{n-1}$ for $0 < n < N$, $h_0 = t_1$, $h_N = 1 - t_N$. A path $Z(0) = i_0 i_1 \dots i_N = Z(1)$ contributes to $\widetilde{f}(x)$ only if $N > 0$ and then by

$$\begin{aligned} &\int_{S_N} \Big[\prod_{n=1}^N \mathrm{e}^{\mu_{i_{n-1}i_{n-1}}h_n} \mu_{i_{n-1}i_n}\mathrm{d}t_n\Big] \cdot \mathrm{e}^{\mu_{i_N i_N}h_N} \cdot \mathbf{1}(b^{i_0}h_1 + \cdots + b^{i_{N-1}}h_N = x) \\ &\qquad \leq \int_{S_N} \prod_{n=1}^N \mu_{i_{n-1}i_n}\mathrm{d}t_n \;\leq\; \overline{\mu}^N \cdot \mathrm{Leb}(S_N) \;=\; \frac{\overline{\mu}^N}{N!} \end{aligned}$$

where $\overline{\mu} = \max \mu_{ij}$. Thus the contribution from all paths of length N is at most $p^N \overline{\mu}^N / N!$. Summing over N gives the bound $\mathrm{e}^{p\overline{\mu}}$ for $f(x)$ which is independent of x. □

Remark 6 The assumption $T < \infty$ does not hold for say whole life insurance or life annuities, but can easily be dispensed with. Consider e.g. whole life insurance where all contributions are paid before time S (say the retirement age) and a lump sum of 1 is paid out at the time τ of the insured's death. Then

$$U(s,T) \;=\; U(s,S)+\mathrm{e}^{-r\tau}1(Z(S)=1)$$

where $\mathrm{e}^{-r\tau}1(Z(S)=1)$ has a bounded density if the mortality rate is piecewise constant (or, in general, is say strictly increasing). Since the density of $U(s,S)$ is bounded according to Theorem 7, similar piecing-together arguments as in the proof there gives the desired conclusion of the density of $U(s,T)$ being bounded.

Similar remarks apply to life annuities where

$$\begin{aligned} U(s,T) &= U(s,S)+1(Z(S)=1)\int_S^\infty \mathrm{e}^{-ru}1(\tau>u\,|\,\tau>S)\,\mathrm{d}u \\ &= U(s,S)+1(Z(S)=1)(1-\mathrm{e}^{-t\tau})/r\,. \end{aligned}$$

9 Numerical examples

We consider two numerical examples serving different purposes. The first illustrates the use of a higher order multi–state Markov process with both rates and lump sums, where the latter in certain cases are only paid with a certain probability. The second example is based on a simpler model from [11] with real (time–varying) rates.

Example 2 We consider a disability–unemployment model as illustrated in Figure 1 (left). For simplicity, and in the light of (2.6), we assume without loss of generality that the parameters are constant over the considered time period. Thus the model is defined in terms of a time–homogeneous Markov process $\{Z(t)\}_{t\geq 0}$, with state space $E=\{1,2,3,4,5\}$, where state 1 corresponds to active, 2 disabled, 3 unemployed, 4 re–employed and 5 death. The parametrisation of the model is thus defined as

$$\boldsymbol{C}=\begin{pmatrix} -0.7 & 0.05 & 0.1 & 0 & 0.5 \\ 0 & -0.5 & 0 & 0 & 0.5 \\ 0 & 0 & -0.7 & 0.1 & 0.5 \\ 0 & 0.05 & 0 & -0.6 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},\quad \boldsymbol{D}=\begin{pmatrix} 0 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix} 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\boldsymbol{b}=(-1,1,1,-1,0).$$

Thus the only lump sum payments (of quantity 2) is upon entering the disability state, 2, from any other (feasible) state. For the red coloured lump sums we see that the corresponding intensities for entering state 2 is split fifty–fifty between the $\boldsymbol{C}$ and $\boldsymbol{D}$ matrices. Thus, only fifty percent of the entrances to state 2 will bring about a lump sum payment of 2. The rationale behind this could e.g. be that disability caused by a work related accident will be indemnified by someone else. From state 3, unemployed, the blue number indicates that work related accident is not a possibility and indemnification will occur with probability one. The active and re–employed states are premium paying states, also at rate 1, and in the unemployed and disability state the insured receives an periodic yearly payment of 1. The interest rate is set to $r = 0.08$, which currently is unrealistically high in most places but reinforces the difference from ordinary moments. The time is set to $T = 10$.

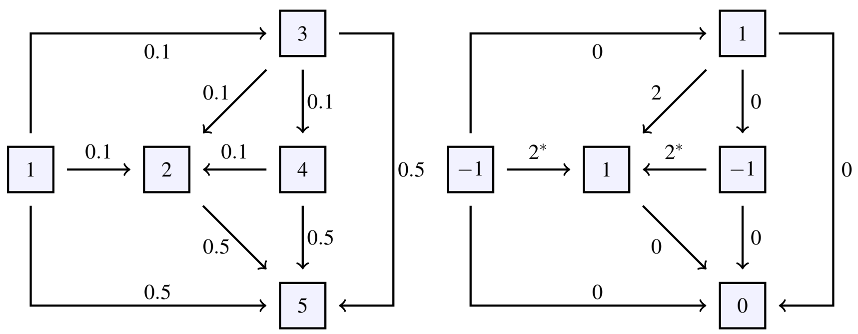


Fig. 1 Disability-Unemployment flow diagram of the underlying Markov process. Left: transition rates. Right: lump sum payments and premium/benefit rates. The lump sums with $*$ are paid with probability of 0.5.

As a first exercise we calculated the first 10 moments of $U(0,T)$ conditionally on $Z(0) = 1$ (active). We compared to simulations of respectively $100,000$, $1,000,000$ and $20,000,000$ realizations from which we calculate the corresponding empirical moments. The results shown in Table 1 indicates that it may be relatively time–consuming to obtain good estimates of higher order moments. The first four moments, which are ingredients used for the calculation of the distributional descriptors such as mean, variance, skewness and kurtosis, are reasonable well estimated even for 100,000 simulations. This is not the case of the fifth moment, where numerical instability due to changing signs makes straightforward averaging of the fifth powers of the data an unstable procedure.

		simulations					
Moment	exact	100,000	Δ	1,000,000	Δ	20,000,000	Δ
1	-0.7248	-0.7326	0.0078	-0.7266	0.0018	0.7247	0.0001
2	3.6404	3.6632	0.0228	3.6458	0.0054	3.6408	0.0004
3	-3.2698	-3.4311	0.1612	-3.3018	0.0320	-3.2700	0.0001
4	56.566	56.5151	0.0510	56.5643	0.0018	56.5771	0.0110
5	-2.9434	-12.0678	9.1243	-4.4342	1.4908	-3.071	0.1275
6	1677.0	1628.53	48.468	1661.72	15.276	1677.31	0.3056
7	2302.3	1509.62	792.71	2164.67	137.66	2289.08	13.250
8	73842	67634	6208.9	72028	1814.6	73838	5.0240
9	223936	149689	74246	209213	14723	222930	1006.7
10	4264367	3582504	681863	4079763	184604	4262636	1731.2
CPU (sec.)	0.006	3.72		33.6		741.2	

Table 1: The first 10 moments. Exact moments vs. simulated moments based on 100,000, 1,000,000 and 20,000,000 repetitions. Δ is the absolute difference to the exact moment.

Next we consider polynomial approximation of the density and distribution function using the moments. Figure 2 shows different approximations with a number of different moments employed. The histogram of the simulated data indicates that the underlying distributions has a challenging shape. It of course requires a very large number of moments to achieve a near perfect fit, and certainly 50 moments may not be enough depending on the purpose of the density. The distribution function, on the other hand, is much easier to approximate, and with 20 moments we get a nearly perfect approximation. From the distribution function we may e.g. obtain the quantiles as an informative and important statistic. The 99% quantile was 4.89 for the 100,000 simulated data while it was 4.95 for all approximations.

Example 3 We consider the following disability-pension model of [11], p.308. A 40 year old male buys an insurance against disability in which case his benefit is a periodic yearly payment of 100,000 DKK until the age of retirement at 65. Whether disabled or not, the insured further receives a pension of 100,000 DKK per year from the age of 65 until his death. We specify a three state Markov model with states 1: active, 2: disabled and 3: dead. The flow diagram is shown in Figure 3. We use the same transition rates as in [11] which are

$$\begin{aligned}
\mu_{12}(x) &= \left(0.0004 + 10^{4.54+0.06x-10}\right) 1_{\{x\leq 65\}} \\
\mu_{21}(x) &= \left(2.0058 \cdot \mathrm{e}^{-0.117x}\right) 1_{\{x\leq 65\}} \\
\mu_{13}(x) &= 0.0005 + 10^{5.88+0.038x-10} \\
\mu_{23}(x) &= \mu_{13}(x)\left(1 + 1_{\{x\leq 65\}}\right).
\end{aligned}$$

We approximate this model by a Markov process which has piecewise constant rates through one year periods and take the rates to be $\mu_{k\ell}(i+0.5)$ for intervals $[i,i+1)$, $i = 40,\ldots$. There are no lump sums involved so $\boldsymbol{B}(t) = \boldsymbol{D}(t) = \boldsymbol{0}$. If $c < 0$ denotes the

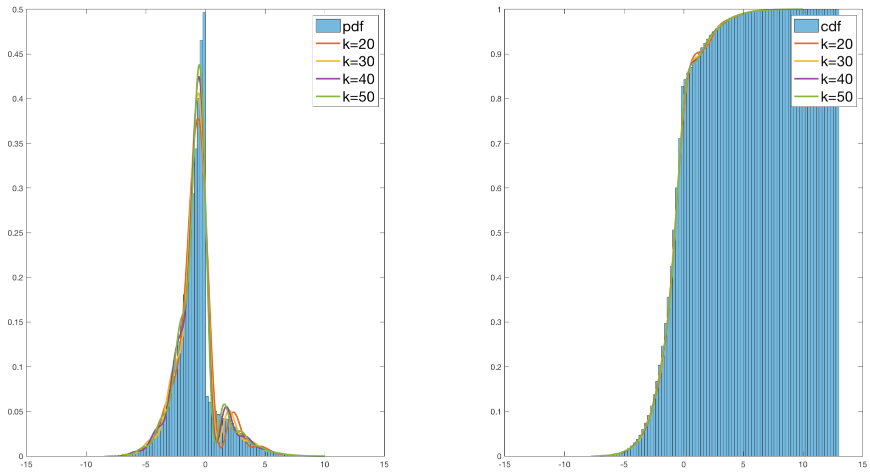


Fig. 2 Left: densities based on $k = 20, 30, 40, 50$ moments plotted towards a histogram based on 100,000 simulations. Right: Empirical c.d.f. of the 100,000 simulations vs. the c.d.f. obtained by k moments

premium rates, then (in units of 100,000 DKK) we have that

$$\boldsymbol{b}(t) = \begin{cases} (c, 1, 0), & t \in [40, 65) \\ (1, 1, 0), & t \geq 65 \end{cases}.$$

The interest rate is set to $r = 0.01$. The fair premium for the continuous model, as reported in [11], is annually $46,409$ DKK to be paid until the age of 65. For our piecewise constant approximation we find the fair premium amount to be $46,419$ DKK annually. Using this amount as the premium rate, i.e. $c = -0.46419$, we calculate the moments up to order 60 in order to approximate the density and distribution functions for the discounted future payments (for justification, see Remark 6). The results are shown in Figure 4. It is clear that the densities are more difficult to approximate than c.d.f.'s, but we see a consistent improvement in the fits with increasing order, and in particular the highest order model using 60 moments seems to capture a number of subtle peculiarities of the underlying distribution as indicated by the histogram. Concerning the c.d.f., all approximations seems to fall very close to each other. Concerning the quantiles, the empirical 99% quantile of the simulated data was 22.18 while the corresponding values for the approximations using $k = 20, 40, 60$ moments are 22.80, 22.55 and 22.55, respectively.

□

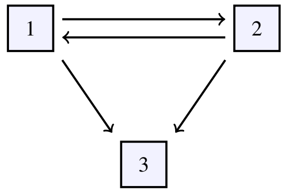


Fig. 3 The disability-pension model from [11]

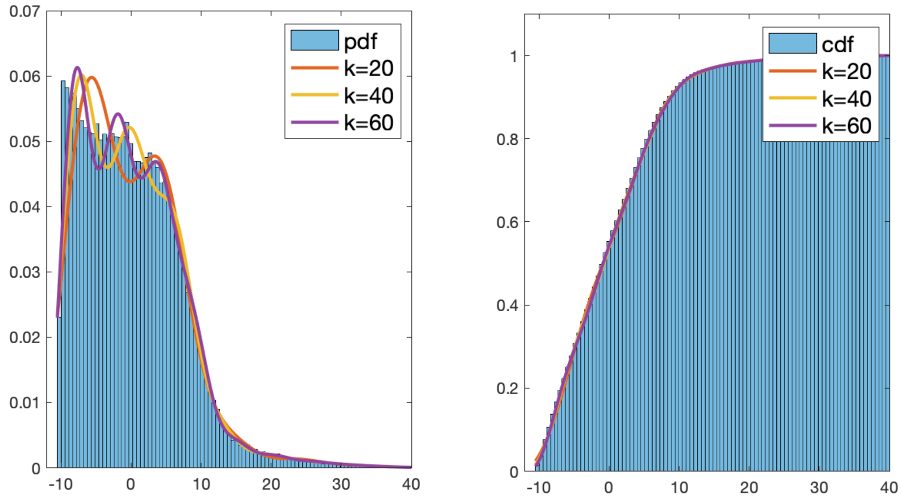


Fig. 4 Left: densities based on $k = 20,40,60$ moments plotted against a histogram based on 100,000 simulations. Right: Empirical c.d.f. of the 100,000 simulations vs. the c.d.f. obtained by the same $k = 20,40,60$ moments.

10 Concluding remarks and future work

In this paper we have established a matrix oriented approach for calculating the (discounted) rewards of time–inhomogeneous Markov processes with finite state–space. In particular, for applications to multi–state Markov models in life insurance our approach provides an alternative to standard derivations in the literature, which are usually based on case–by–case derivations involving differential or integral equations. In the slightly more general set–up in this paper, dealing with partial reserves, we provide a unifying approach to deriving reserves and moments (of, in principle, arbitrary orders) which has a simple numerical implementation. The moment generating function of the (discounted) future payments, which plays an important role in the derivation of the moments and whose derivation is based on probabilistic (sample path) arguments, has a strikingly simple form which would allow for a numerical inversion in order to obtain the c.d.f. or density of the future payments as well. However, since the moments of all orders, in principle, are available we propose an alternative method involving approximation of the c.d.f. and densities via orthogonal polynomial expansions based on central moments.

While this method seems to be very robust concerning the c.d.f., the approximation of the density itself is more involved which stems from the fact that presence of lump sums mixed with continuous rates implies that the densities can have a very challenging form. This strongly suggest to use a reference density f_0 different from the normal and this is in fact necessary in life insurance models allowing repeated lump sum payments at transitions. Note that f_0 need not be chosen to be of a particularly simple form — the orthonormal polynomials are easily available via Gram-Schmidt orthogonalization provided only the moments are readily available, cf. [12].

As mentioned in the Introduction, Markovian point processes provide one more application area of our methodology. The details have been deferred to a future paper, but we note at this place that also here it is necessary to work with a non-normal f_0.

It may also be mentioned that our Markov reward model may be seen as an inhomogeneous Markov additive process (cf. [3], Section XI.1). This leaves another open problem since we do not achieve full generality from this point of view because jumps (lump sums) are not allowed to have random sizes.

The partial reserves are here introduced and defined as a computing efficient ingredient in calculating e.g. the real reserve as their sum. However, there may be other uses of these quantities, in particular, the reserve, conditional on the terminal state of the insured as suggested in (5.3). [25] discussed aspects of the present value of future payments beyond what goes into the balance scheme. [25] mentioned specifically the interest in calculation of policy holder path dependent present values, e.g. in connection with prognostication or more general reporting to policy holders. In that connection the quantity can help answering questions of the type: What do I get out of my contract if I survive until retirement? The point is that, in contrast to the balance scheme, the individual policy holder is not necessarily interested in information based on the expected future policy holder path but interested in as-if or scenario calculations based on a (partially) given future policy holder path. Generalizations of the partial reserves introduced here include reserves conditional on being in a certain series of states at a certain series of future time points or conditional on uninterrupted sojourn in a given state in a future time period. They make it is possible to answer questions of the type (knock on wood): What do I get out of my contract if I manage to hold on to both my good health (as opposed to becoming disabled), my job (as opposed to becoming unemployed), and my marriage (as opposed to becoming divorced or a widower) until retirement, perhaps even relative to losing either of the three? Maybe, even the involvement of corresponding higher order moments are relevant to an (educated) policy holder. Our introduction of partial reserves and moments is just the starting point. Potential applications in policyholder communication is clear. It is left for future research to introduce quantities, calculations, and examples of interest within that specific area.

Many contracts involve lump sum payments as deterministic times such as the retirement age. For notational convenience, we have omitted this feature in the description of the payment stream. Computationally it is, however, easily handled by using our matrix calculations between such deterministic times, adding lump sums and piecing together.

References

1. Abate, J., Whitt, W.: Numerical inversion of Laplace transforms of probability distributions. INFORMS Journal on Computing **7**, 36–43 (1995)
2. Albrecher, H., Teugels, J.L., Tichy, R.F.: On a gamma series expansion for the time-dependent probability of collective ruin. Insurance: Mathematics and Economics **29**(3), 345 – 355 (2001)
3. Asmussen, S.: Applied Probability and Queues. Springer, New York, 2nd Edition (2003)
4. Asmussen, S., Goffard, P.O., Laub, P.J.: Orthonormal polynomial expansions and lognormal sum densities. In: P. Barrieu (ed.) Ragnar Norberg. Risk and Stochastics, chap. 6, pp. 127–150 (2019)
5. Asmussen, S., Koole, G.: Marked point processes as limits of Markovian arrival streams. Journal of Applied Probability **30**(2), 365–372 (1993)
6. Asmussen, S., Steffensen, M.: Risk and Insurance. A Graduate Text. Probability theory and Stochastic Modelling. Springer Verlag (to appear)
7. Barndorff-Nielsen, O., Cox, D.: Asymptotic Techniques for Use in Statistics. Monographs on Statistics and Applied Probability. Springer US (2013)
8. Bladt, M., Meini, B., Neuts, M., Sericola, B.: Distributions of reward functions on continuous-time Markov chains. In: 4th International Conference on Matrix Analytic Methods in Stochastic Models (MAM4). Adelaide, Australia (2002)
9. Bladt, M., Nielsen, B.F.: Matrix-Exponential Distributions in Applied Probability. Springer Verlag (2017)
10. Bowers, N.: Expansion of probability density functions as a sum of gamma densities with applications in risk theory. Trans. Soc. Actuar. **18**(52), 125 – 137 (1966)
11. Buchardt, K., Møller, T.: Life insurance cash flows with policyholder behavior. Risks **3**(3), 290–317 (2015)
12. Cramér, H.: Mathematical Methods of Statistics. (Almqvist & Wiksells Akademiska Handböcker). Princeton University Press (1946)
13. Dobrushin, R.L.: Generalization of Kolmogorov's equations for Markov processes with a finite number of possible states. Mat. Sb. (N.S.) **33(75)**, 567–596 (1953)
14. Gill, R.D., Johansen, S.: A survey of product-integration with a view toward application in survival analysis. The Annals of Statistics **18**(4), 1501–1555 (1990)
15. Goffard, P.O., Laub, P.J.: Two numerical methods to evaluate stop-loss premiums. Journal of Computational and Applied Mathematics (2019). To appear
16. Goffard, P.O., Loisel, S., Pommeret, D.: A polynomial expansion to approximate the ultimate ruin probability in the compound poisson ruin model. Journal of Computational and Applied Mathematics **296**, 499 – 511 (2016)
17. Goffard, P.O., Loisel, S., Pommeret, D.: Polynomial approximations for bivariate aggregate claims amount probability distributions. Methodology and Computing in Applied Probability **19**(1), 151–174 (2017)
18. Hesselager, O., Norberg, R.: On probability distributions of present values in life insurance. Insurance: Mathematics and Economics **18**(1), 35 – 42 (1996)
19. Hoem, J.M.: Markov chain models in life insurance. Blätter der DGVFM **9**(2), 91–107 (1969)
20. Johansen, S.: Product integrals and Markov processes. CWI Newsletter **12**, 3–13 (1986)
21. Loan, C.v., Moler, C.: Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Review **45**, 1–46 (2003)
22. Moler, C., Loan, C.v.: Nineteen dubious ways to compute the exponential of a matrix. SIAM Review **20**(4), 801–836 (1978)
23. Norberg, R.: Reserves in life and pension insurance. Scandinavian Actuarial Journal **1991**(1), 3–24 (1991)
24. Norberg, R.: Differential equations for moments of present values in life insurance. Insurance: Mathematics and Economics **17**(2), 171 – 180 (1995)

25. Norberg, R.: On bonus and bonus prognoses in life insurance. Scandinavian Actuarial Journal **2001**(2), 126–147 (2001)
26. Protter, M., Morrey, C.: Intermediate Calculus. Undergraduate Texts in Mathematics. Springer New York (1996)
27. Ramlau-Hansen, H.: Hattendorff's theorem: A Markov chain and counting process approach. Scandinavian Actuarial Journal **1988**(1-3), 143–156 (1988)
28. Szegö, G.: Orthogonal Polynomials. American Mathematical Society Colloquium Publications, 23. Providence, RI: American Mathematical Society (1939)
29. Van Loan, C.: Computing integrals involving the matrix exponential. IEEE Transactions on Automatic Control **23**(3), 395–404 (1978)

A On the L_2 condition without piecewise constancy

Remark 7 Condition (iii) in Theorem 7 is not far from being necessary. Consider as an example the disability model in the time interval $[0,1]$ with states 0: active, 1: disabled, 2: dead and allowing recovery, with the same intensity $\lambda > 0$ for transitions from 0 to 1 as from 1 to 0 and (for simplicity) mortality rate 0 and discounting rate $r = 0$. The benefits are a lump sum $b^{01} = 1$ upon transition from 0 to 1 and the contributions a constant payment rate $b^0 < 0$ when active.

When $Z(0) = 0$, the total number N of transitions in $[0,1]$ is Poisson(λ), the total benefits $U_1(0,1)$ are $M = \lceil N/2 \rceil$, and the total contributions $U_0(0,1)$ equal $|b^1|$ times the total time T_0 spent in state 0. Thus $U(0,1) = U_1(0,1) - U_0(0,1) = M - U_0(0,1)$ Obviously, $U_0(0,1)$ is concentrated on the interval $[0,|b^1|]$ with a density $g(x)$ which is bounded away from 0, say $h(x) > c_1 > 0$.

Assuming, again for simplicity, that $b^1 > 1$. Then the intervals $[m - b^1, m]$ overlap and so for a given x, at least one of them contribute to $f(x)$. One candidate is the one with $m = \lceil x \rceil]$. This gives

$$f(x) \;\geq\; \mathbb{P}(M = \lceil x \rceil) h(\lceil x \rceil - x) \;\geq\; \mathbb{P}(N = 2\lceil x \rceil) c_1 \,.$$

But using Stirling's approximation to estimate the Poisson probability, it follows after a little calculus that $\mathbb{P}(N = 2\lceil x \rceil) \geq c\mathrm{e}^{-4x\log x}$ for all large x, say $x \geq x_0$, and some $c > 0$ (in fact the 4 can be replaced by any $c_3 > 2$). This gives for a normal(μ,σ^2) f_0 that

$$\int \frac{f^2}{f_0} \;\geq\; \int_{x_0}^{\infty} \frac{f^2(x)}{f_0(x)}\,\mathrm{d}x \;\geq\; c_4 \int_{x_0}^{\infty} \exp\{(x-\mu)/\sigma^2 - 8x\log x\}\,\mathrm{d}x \;=\; \infty .$$

That is, (8.1) fails and a similar calculation gives that (8.5) does so too.

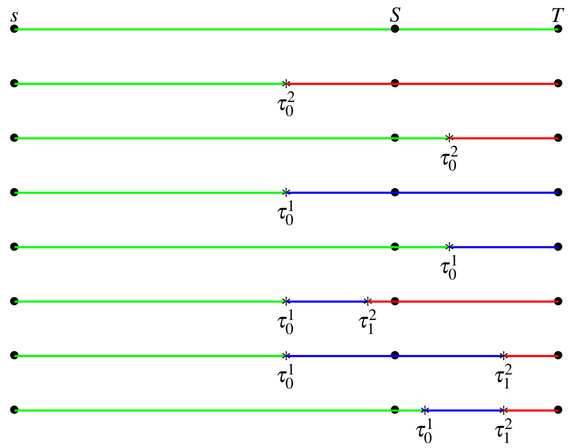
The obvious way out is of course to take f_0 with a heavier tail than the normal, say doubly exponential (Laplacian) with density $\mathrm{e}^{-|x|}$ for $-\infty < x < \infty$. Given that this example and other cases where condition (iii) is violated do not seem very realistic, we have not exploited this further.

Example 4 Consider again the disability model with states 0: active, 1: disabled, 2: dead and no recovery. As in [11], we assume that the payment stream has the form $b^0(t) = -b^0_-$ for $t \leq S$ and $b^0(t) = b^0_+$ for $t > S$,[1] $b^1(t) \equiv b^1$ for all t. For example S could be the retirement age, say 65. Without recovery, the only non-zero transition rates are the $\mu_{01}(t)$, $\mu_{02}(t)$, $\mu_{12}(t)$. With the values used in [11], these are bounded away from 0 and ∞ on $[0,T]$ for any T (say 75 or 80), and this innocent assumption is all that matters for the following.

Define the stopping times

$$\begin{aligned}
\tau_0^1 &= \inf\{t > s : Z(t) = 1, Z(s) = 0 \text{ for } s < t\} \\
\tau_1^2 &= \inf\{t > \tau_0^1 : Z(t) = 2\} \\
\tau_0^2 &= \inf\{t > s : Z(t) = 2, Z(s) = 0 \text{ for } s < t\}
\end{aligned}$$

[1] Here and in the following, a - subscript mimics 'before S' and a + 'after S'.



$F_0 = \{\tau_0^1 \wedge \tau_0^2 > T\}$

$F_{2-} = \{0 < \tau_0^2 \le S\}$

$F_{2+} = \{S < \tau_0^2 \le T\}$

$F_{1-} = \{0 < \tau_0^1 \le S,\, \tau_1^2 > T\}$

$F_{1+} = \{S < \tau_0^1,\, \tau_1^2 > T\}$

$F_{1_-2_-} = \{0 < \tau_0^1 < \tau_1^2 \le S\}$

$F_{1_-2_+} = \{0 < \tau_0^1 \le S < \tau_1^2 \le T\}$

$F_{1_+2_+} = \{0 < \tau_0^1 \le S < \tau_1^2 \le T\}$

Fig. 5 The sample space partition. State 0=active=green, State 1=disabled=blue, State 2=dead=red

with the usual convention that the stopping time is ∞ if there is no t meeting the requirement in the definition. One then easily checks that the sets $F_0,\dots,F_{1_+2_+}$ defined in Fig. 5 defines a partition of the sample space. Here F_0 (corresponding to zero transitions in $[s,T]$) contributes with an atom at

$$a = -b_-^0(1-\mathrm{e}^{-rS})/r + b_+^0(\mathrm{e}^{-rS}-\mathrm{e}^{-rT})/r \ \text{ with probability } \ q = \exp\Big\{\int_0^T \mu_{00}(u)\,\mathrm{d}u\Big\}$$

and the remaining 7 events with absolutely continuous parts, say with (defective) densities $g_{2_-},\dots,g_{1_+2_+}$. It is therefore sufficient to show that each of these is bounded. In obvious notation, the contribution to $U(s,T)$ of the first 4 of these 7 events (corresponding to precisely one transition in $[s,T]$) are

$$\begin{aligned} A_{2-} &= \big[-b_-^0(1-\mathrm{e}^{-r\tau_0^2})/r\big]\cdot 1_{\tau_0^2\le S},\\ A_{2-} &= \big[-b_-^0(1-\mathrm{e}^{-rS})/r + b_+^0(\mathrm{e}^{-rS}-\mathrm{e}^{-r\tau_0^2})/r\big]\cdot 1_{S<\tau_0^2\le T},\\ A_{1-} &= \big[-b_-^0(1-\mathrm{e}^{-r\tau_0^1})/r + b^1(\mathrm{e}^{-rT}-\mathrm{e}^{-r\tau_0^1})/r\big]\cdot 1_{\tau_0^1\le S},\\ A_{1_+} &= \big[-b_-^0(1-\mathrm{e}^{-rS})/r + b_+^0(\mathrm{e}^{-rS}-\mathrm{e}^{-r\tau_0^1})/r + b^1(\mathrm{e}^{-rT}-\mathrm{e}^{-r\tau_0^1})/r\big]\cdot 1_{S<\tau_0^1\le T}. \end{aligned}$$

The desired boundedness of $g_{2-}, g_{2+}, g_{1-}, g_{1_+}$ therefore follows from the following lemma, where τ may be improper so that $\int h < 1$:

Lemma 3 *If a r.v. τ has a bounded density h and the function φ is monotone and differentiable with φ' bounded away from 0, then the density of $\varphi(\tau)$ is bounded as well.*

Proof The assumptions imply the existence of $\psi = \varphi^{-1}$. Now just note that the density of $\varphi(\tau)$ is $h(\psi(x))/\varphi'(\psi(x))$ if φ is increasing and $h(\psi(x))/\big|\varphi'(\psi(x))\big|$ if it is decreasing. □

The cases of $g_{1_-2_-}, g_{1_-2_+}, g_{1_+2_+}$. (corresponding to precisely two transitions in $(s,T]$) is slightly more intricate. Consider first $g_{1_-2_-}$. The contribution to $U(0,T)$ is here

$$A_{1_-2_-} = -b_-^0\int_0^{\tau_0^1}\mathrm{e}^{-ru}\,\mathrm{d}u + b^1\int_{\tau_0^1}^{\tau_1^2}\mathrm{e}^{-ru}\,\mathrm{d}u = c_0 + c_1\mathrm{e}^{-r\tau_0^1} - c_2\mathrm{e}^{-r\tau_1^2}$$

Now the joint density $h(t_1,t_2)$ of (τ_0^1,τ_0^2) at a point (t_1,t_2) with $0<t_1<t_2\le R$ is

$$\exp\Big\{\int_0^{t_1}\mu_{00}(u)\,\mathrm{d}u\Big\}\mu_{01}(t_1)\cdot\exp\Big\{\int_{t_1}^{t_2}\mu_{11}(v)\,\mathrm{d}v\Big\}\mu_{12}(t_2)$$

so that $h(t_1,t_2)$ is bounded. Consider now the transformation taking (τ_0^1,τ_0^2) into (τ_0^1,A_{12-}). The inverse of the Jacobiant J is

$$\begin{vmatrix} 1 & 0 \\ -rc_1\mathrm{e}^{-r\tau_0^1} & rc_2\mathrm{e}^{-r\tau_1^2} \end{vmatrix} = rc_2\mathrm{e}^{-r\tau_1^2}$$

which is uniformly bounded away from 0 when $\tau_1^2 \le S$. Therefore also the joint density $k(z_1,z_2)$ of (τ_0^1,A_{12-}) is bounded. Integrating $k(z_1,z_2)$w.r.t. z_1 over the finite region $0 \le z_1 \le S$ finally gives that $g_{1_-2_-}$ is bounded.

The argument did not use that $\tau_1^2 \le S$ and hence also applies to $g_{1_-2_+}$. Finally note that the contribution to $g_{1_+2_+}$ from $[0,S]$ is just a constant, whereas the one from $(S,T]$ has the same structure as used for $g_{1_-2_-}$. Hence also $g_{1_+2_+}$ is bounded.

Remark 8 The calculations show that F_{1+} also is an atom if $b_+^0 = b^1$. However, since the contribution rate b_-^0 is calculated via the equivalence principle given the benefits b_+^0, b^1 and the transition intensities, this would be a very special case. The somewhat less special situation $b_+^0 = b^1$ (same annuity to an disabled as to someone retired as active) also gives an atom, now at F_{1+}. In fact, $b_-^0 = b^1$ is assumed in [11].

Remark 9 For a counterexample to (8.1), consider again the disability model with the only benefit being a lump sum of size $b^{01}(t) = \mathrm{e}^{rt}\varphi(t)$ being paid out at τ_0^1 where $\varphi(t) = (t-a)^2 1_{t\le a} + b$, cf. Fig. 6.

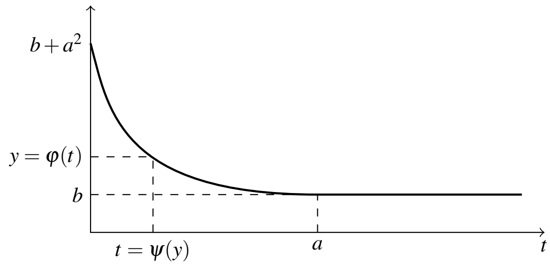


Fig. 6 $\varphi(t)$

Then $U(s,T] = \varphi(\tau_0^1)$ has an atom at b (corresponding to $\tau_0^1 > a$) and an absolutely continuous part on $(b, b+a^2]$. Letting $\psi : [b,b+a^2] \mapsto [0,a]$ be the inverse of φ, $t = \psi(y)$ satisfies

$$y = (t-a)^2 + b \ \Rightarrow (y-b)^{1/2} = -(t-a) \ \Rightarrow \ t = \psi(y) = a - (y-b)^{1/2},$$

where the minus sign after the first $\Rightarrow$ follows since ψ is decreasing. With h the density of τ_0^1, we get as in Lemma 3 that the conditional density f of the absolutely continuous part is given by

$$f(y) = \frac{1}{\mathbb{P}(\tau_0^1 \le a)} \frac{h(\psi(y))}{|\varphi'(\psi(y))|} = \frac{1}{\mathbb{P}(\tau_0^1 \le a)} \frac{h(a-(y-b)^{1/2})}{(y-b)^{1/2}}$$

But there are $c_1, c_2 > 0$ such that $f_0(y) \le c_1$ on $[b,b+a^2]$ and $h(t) \ge c_2$ on $[b,b+a^2]$, and so we get

$$\int \frac{f^2}{f_0} = \int_b^{b+a^2} \frac{f^2(y)}{f_0(y)}\,\mathrm{d}y \ \ge \ \int_b^{b+a^2} \frac{c_2^2}{c_1(y-b)}\,\mathrm{d}y = \infty,$$

meaning that (8.1) does not hold.