Continuous Affine Processes: Transformations, Markov Chains and Life Insurance

Kristian Buchardt*
University of Copenhagen and PFA Pension

October 27, 2014

Abstract

Affine processes possess the property that expectations of exponential affine transformations are given by a set of Riccati differential equations, which is the main feature of this popular class of processes. In this paper we generalise these results for expectations of more general transformations. This is of interest in e.g. doubly stochastic Markov models, in particular in life insurance. When using affine processes for modelling the transition rates and interest rate, the results presented allow for easy calculation of transition probabilities and expected present values.

Keywords: affine processes; doubly stochastic process; multi-state life insurance models; credit risk; surrender modelling; stochastic mortality; stochastic interest

2010 Mathematics Subject Classification: Primary: 97M30 Secondary: 60J60;91G40;91G30

1 Introduction

The main results of this paper is a generalisation of a result from [5] which provides differential equations for calculating the expectation of a certain transformation of affine processes. We present a, to the author's knowledge, new proof of the result, which is constructive and allows for generalisation to expectations of more general transformations. These results are interesting as a part of the mathematical study of the class of affine processes. The results are also applicable, and the class of transformations are useful for the use of affine processes as transition rates in doubly stochastic Markov

*Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen O and PFA Pension, Sundkrogsgade 4, DK-2100 Copenhagen O, Denmark, http://math.ku.dk/~buchardt, e-mail: buchardt@math.ku.dk

chains. Such Markov chains are used in e.g. life insurance mathematics and credit risk modelling, where it is also possible to model e.g. the interest rate jointly as an affine process. In life insurance, one of the main contributions from these results is the ability to handle dependent interest and transition rates while exploiting the structure of affine processes.

In traditional life insurance mathematics, a finite state Markov chain is often chosen to represent the state of the insured. Associating payments with sojourns in states and transitions between states, one can easily find the expected present value of life insurance liabilities when an interest rate is given. Traditionally these models have been studied with a deterministic interest rate and deterministic transition rates, however in recent life insurance mathematics, stochastic modelling of the interest and transition rates has gained attention. This is of particular interest either if one wants to study the risk associated with changes in the underlying interest and transition rates, and/or if one wants to hedge this risk in securities based on these same underlying rates. A basic reference on the life insurance setup with stochastic transition rates is [11], where the underlying rates are modelled by a finite state Markov chain. In particular, dependence between the rates is allowed. Basic treatment of stochastic interest rates applied in life insurance is given in [10], and for stochastic mortality rates see e.g. [3]. For combined models for stochastic interest and mortality rates see [4] and [1]. Common for several of the references mentioned is that interest and mortality rates are modelled as affine processes. This class of processes leads to mathematically tractable models, where one can solve a system of ordinary differential equations instead of partial differential equations in order to find expected present values. The results presented in this paper allow for the application of affine processes in more general life insurance models, thus making it easier to calculate expected present values.

Finite state Markov chains are also used when modelling credit risk, see e.g. [7]. A basic credit risk model is a two state Markov chain, where a jump from the initial state represents a default. A popular extension of this model is to let the default transition rate be modelled as a stochastic process itself such that it is possible for it to be dependent on the interest rate and other economic factors. This approach is studied in [9] where various Markov chain models are considered. A more general treatment of the Markov chain approach to credit risk modelling with stochastic transition rates is given in [8], and it is shown how prices generally satisfy a system of partial differential equations. In both papers, it is shown how one can benefit from affine stochastic processes as transition intensities and economic factors. If the model is particularly simple, the Riccati equations and a result from [5] can be used to reduce the problem of solving a system of partial differential equations to that of solving a system of ordinary differential equations. The results presented in this paper generalize these methods. This allows us to find prices in more general decrement Markov chain models solving only ordinary differential equations instead of partial differential equations.

The paper is structured as follows. In Section 2 we introduce the basics of affine processes that are multidimensional, continuous and time-inhomogeneous, allowing us to state the main results precisely. From this we present the main results: We give a new proof for a result from [5] on the expectation of a transformation of affine processes and use this to formulate and prove a more general result on affine processes. The results are discussed and further generalisations are considered. These results can be applied for doubly stochastic Markov chains, in particular in life insurance and credit risk modelling. In Section 3 we show how the results can be applied for calculation of transition probabilities in certain doubly stochastic Markov chains, and we present an example, which is studied numerically as well. In Section 4 we show how to apply the results for valuation of life insurance contracts.

2 Continuous affine processes

In this section we study continuous affine processes, inspired by Chapter 10 in [6]. In Section 2.1 we give a proper definition and then present Theorem 2.2, which is the usual theorem and basic property about affine processes, adapted to our setup. This gives e.g. bond prices and survival probabilities in the case where the short rate, respectively the mortality rate, is a stochastic affine process. This provides the basis for the main results which are presented in Section 2.2.

2.1 Definition and set-up

Let $(\Omega, \mathcal{F}, (\mathcal{F}(t))_{t\in\mathbb{R}_+}, P)$ be a filtered probability space, satisfying the usual conditions, and denote by $\mathbf{W} = (W(t))_{t\in\mathbb{R}_+}$ an adapted d-dimensional Wiener process. Let $\mathbf{X} = (X(t))_{t\in\mathbb{R}_+}$ be a d-dimensional stochastic process satisfying the stochastic differential equation,

$$\mathrm{d}X(t) = \delta(t, X(t))\,\mathrm{d}t + \rho(t, X(t))\,\mathrm{d}W(t), \tag{2.1}$$

with $X(0) = x \in \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^d$ is the state space. The functions $\delta : \mathbb{R}_+ \times \mathcal{X} \to \mathbb{R}^d$ and $\rho : \mathbb{R}_+ \times \mathcal{X} \to \mathbb{R}^{d\times d}$ are assumed to be measurable. We have not assumed any Lipschitz continuity of the parameter functions δ and ρ, and we allow for some discontinuities. In this paper we make the crucial assumption that $\mathbf{X}$ exists for all start values $x \in \mathcal{X}$, and start time points.

A continuous stochastic process $\mathbf{X}$ is affine on $[0, T]$ if the $\mathcal{F}(t)$-conditional characteristic function of $X(T)$ has an exponential affine form, and this is made precise in the following definition. We think of T as a fixed, long time horizon.

Definition 2.1. *The process* $\mathbf{X}$*, with initial value* $X(0) = x$*, is* affine *on* $[0, T]$ *if there exist functions* ϕ *and* ψ *such that for all* $x \in \mathcal{X}$*,* $0 \le t \le T$ *and* $z \in \mathbb{R}^d$*,*

$$\mathrm{E}\left[e^{\mathrm{i}z^\top X(T)} \,\middle|\, \mathcal{F}(t) \right] = e^{\phi(t,T,z) + \psi(t,T,z)^\top X(t)} \tag{2.2}$$

holds, where $\phi(t, T, z)$ *is* $\mathbb{C}$*-valued and* $\psi(t, T, z)$ *is* $\mathbb{C}^d$*-valued.*

It can be shown that for $\mathbf{X}$ to be affine, it is a necessary condition that the drift and diffusion parameter functions are affine of the form

$$\begin{aligned} \delta(t, x) &= b(t) + \sum_{i=1}^{d} \beta_i(t) x_i = b(t) + \mathcal{B}(t) x, \\ \rho(t, x)\rho(t, x)^\top &= a(t) + \sum_{i=1}^{d} \alpha_i(t) x_i, \end{aligned} \tag{2.3}$$

for some vector functions $b : \mathbb{R}_+ \to \mathbb{R}^d$ and $\beta_i : \mathbb{R}_+ \to \mathbb{R}^d$, $i = 1, \ldots, d$, and matrix functions $a : \mathbb{R}_+ \to \mathbb{R}^{d\times d}$ and $\alpha_i : \mathbb{R}_+ \to \mathbb{R}^{d\times d}$, $i = 1, \ldots, d$. We have $\mathcal{B}(t) = (\beta_1(t), \ldots, \beta_d(t))$, i.e. column i equals $\beta_i(t)$. We use the notation $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \ge 0\}$.

From now on we assume that (2.3) holds, and that the parameter functions b, β_i, a, α_i, $i = 1, \ldots, i$ are bounded and piecewise continuous.

In order to allow for modelling flexibility, we consider affine transformations of $\mathbf{X}$. Let $p \ge 1$, and let c and Γ be a vector and matrix function respectively, $c : \mathbb{R}_+ \to \mathbb{R}^p$ and $\Gamma : \mathbb{R}_+ \to \mathbb{R}^{p\times d}$. Define the p-dimensional process $\mathbf{Y}$ by

$$Y(t) = c(t) + \Gamma(t) X(t). \tag{2.4}$$

We assume that c_j and Γ_{ji}, $j = 1, \ldots, p$, $i = 1, \ldots, d$ are piecewise continuous and bounded, with limits everywhere. It is noted, that if $\Gamma(t)$ has a left inverse for all t, or, equivalently if it is injective for all t, then the affine transformation $\mathbf{Y}$ is an affine process on some state-space, as can be seen by the definition (2.2).

We use the column sums of Γ, so define the d-dimensional function $\gamma(t) = \mathbb{1}^\top \Gamma(t)$, where $\mathbb{1} = (1, \ldots, 1)^\top$ is a column vector with 1 in all entries. Then $\gamma_i(t) = \mathbb{1}^\top \Gamma(t) e_i$ is the sum of column i in Γ, where e_i is the ith standard unit vector. Using this notation, we have $\mathbb{1}^\top Y(t) = \mathbb{1}^\top c(t) + \sum_{i=1}^{d} \gamma_i(t) X_i(t)$.

The following theorem states the essential feature of the affine processes, and the result is the reason for the great interest in affine processes.

Theorem 2.2. *Let* $\mathbf{X}$ *be an affine process. If* (ϕ, ψ) *exists as a solution to* (2.7) *and the stochastic process*

$$t \mapsto e^{-\int_0^t \mathbb{1}^\top Y(s)\,\mathrm{d}s + \phi(t,T) + \psi(t,T)^\top X(t)} \tag{2.5}$$

is a martingale, then

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] = e^{\phi(t,T)+\psi(t,T)^\top X(t)}, \tag{2.6}$$

where the functions ϕ and ψ solve the Riccati differential equations,

$$\begin{aligned} \frac{\partial}{\partial t}\phi(t,T) &= -\frac{1}{2}\psi(t,T)^\top a(t)\psi(t,T) - b(t)^\top \psi(t,T) + \mathbb{1}^\top c(t), \\ \phi(T,T) &= 0, \\ \frac{\partial}{\partial t}\psi_i(t,T) &= -\frac{1}{2}\psi(t,T)^\top \alpha_i(t)\psi(t,T) - \beta_i(t)^\top \psi(t,T) + \gamma_i(t), \quad i = 1,\dots,d, \\ \psi(T,T) &= 0. \end{aligned} \tag{2.7}$$

For the case of time-homogeneous parameters, the theorem is proven in [6], and this can be generalized to the case of time-inhomogeneous piece-wise continuous parameter functions presented here.

We briefly consider some important applications of Theorem 2.2. Consider the case where $\mathbf{X}$ is two-dimensional and jointly models the short rate and mortality rate, i.e. $X(t) = (X_1(t), X_2(t)) = (r(t), \mu(t))$. Choosing

$$\Gamma(t) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and $c(t) = 0$, we obtain by (2.4) that $Y^r(t) = r(t)$, i.e. $\mathbf{Y}^r$ is the short rate. If we work under a pricing measure, Theorem 2.2 can be applied to obtain bond prices,

$$P(t,T) = \mathrm{E}\left[e^{-\int_t^T Y^r(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] = e^{\Phi^r(t,T)+\Psi^r(t,T)^\top X(t)},$$

where $P(t,T)$ is the market value at time t of a payment of 1 at time T. Similarly, by choosing another Γ, we can let $\mathbf{Y}^\mu$ be the mortality rate, i.e. $Y^\mu(t) = \mu(t)$. Then, if we work under the physical measure, the survival probabilities are obtained,

$$S(t,T) = \mathrm{E}\left[e^{-\int_t^T Y^\mu(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] = e^{\Phi^\mu(t,T)+\Psi^\mu(t,T)^\top X(t)},$$

where $S(t,T)$ is the probability of surviving from time t till time T. Finally, choosing Γ as the identity matrix, $\mathbf{Y}$ is two-dimensional and equal to $\mathbf{X}$, hence jointly modelling the short rate and the mortality rate. In this case, if we work under a pricing measure, we obtain the price of a pure endowment,

$$V(t) = \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] = e^{\Phi(t,T)+\Psi(t,T)^\top X(t)},$$

where $V(t)$ is the expected present value of a payment of 1 at time T, conditional on survival to time T given that the individual is alive at time t. Note that here the interest and mortality rate can be dependent. If we consider the special case of independence between the interest and mortality rate, we obtain $V(t) = P(t,T)S(t,T)$ and in particular that

$$\Phi(t,T) = \Phi^r(t,T) + \Phi^\mu(t,T),$$
$$\Psi(t,T) = \Psi^r(t,T) + \Psi^\mu(t,T).$$

For an endowment insurance, where there is a payment upon death as well, Theorem 2.2 is not sufficient for pricing, and we have to use Theorem 2.3 below.

2.2 Main results

In this section we present the main contributions of the paper, which is basically a new proof of Theorem 2.3 below. The advantage of the proof is that it is constructive and the idea of the proof can be reapplied which allows us to state and prove Theorem 2.7, which, to the author's knowledge, is a new result. Together with Theorem 2.2 the two theorems presented in this section have applications for Markov chain modelling with stochastic transition rates, e.g. for life insurance valuation or credit risk modelling.

Consider the transformation

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) \middle| \mathcal{F}(t) \right],$$

for $k \in \{1, \dots, p\}$ and $u \in [t,T]$, and recall that $\mathbf{Y}$ is defined by (2.4). In [5] differential equations are derived for the expectation under slightly different conditions for time-homogeneous affine jump diffusions, though only for the case where $u = T$. The result presented here is for a general $u \in [t,T]$ and the case of continuous affine processes with time-inhomogeneous coefficients. The system of differential equations found in [5] is essentially the same as the one presented in Theorem 2.3.

Theorem 2.3. *Let $k \in \{1, \dots, p\}$ and $u \in [t,T]$. Then, if either Assumption 2.4 or Assumption 2.5 holds, and under the conditions of Theorem 2.2, it holds that*

$$\begin{aligned} &\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) \middle| \mathcal{F}(t) \right] \\ &= e^{\phi(t,T) + \psi(t,T)^\top X(t)} \left(A^k(t,T,u) + B^k(t,T,u)^\top X(t) \right), \end{aligned} \tag{2.8}$$

where (ϕ, ψ) is a solution to (2.7) and (A^k, B^k) solves the linear differential equation *system,*

$$\begin{aligned}
\frac{\partial}{\partial t}A^k(t,T,u) &= -\psi(t,T)^\top a(t)B^k(t,T,u) - b(t)^\top B^k(t,T,u),\\
A^k(u,T,u) &= e_k^\top c(u),\\
\frac{\partial}{\partial t}B_i^k(t,T,u) &= -\psi(t,T)^\top \alpha_i(t)B^k(t,T,u) - \beta_i(t)^\top B^k(t,T,u), \quad i=1,\dots,d\\
B^k(u,T,u) &= e_k^\top \Gamma(u).
\end{aligned} \tag{2.9}$$

The original proof of the result, from [5], is the classic one, which holds under the following assumption.

Assumption 2.4. *(Classic integrability condition) The process*

$$t \mapsto e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s + \phi(t,T) + \psi(t,T)^\top X(t)} \left(A^k(t,T,u) + B^k(t,T,u)^\top X(t)\right)$$

is a real martingale (i.e. not only a local martingale).

The assumption is that, when multiplied with $e^{-\int_0^t \mathbb{1}^\top Y(s)\,\mathrm{d}s}$, the right hand side of (2.8) is a martingale. Then, by an application of Itô's lemma to the martingale, and by setting the drift equal to zero, one can obtain the system of differential equations. Carrying out this proof, note that $A^k(t,T,u)$ and $B^k(t,T,u)$ are constant for $t > u$.

Instead of presenting the original proof, we give a new proof. Below is presented an outline of the proof, where certain details are omitted. The result holds under a slightly different integrability condition than in the original proof. It holds whenever we can interchange differentiation and expectation, and this is allowed if we can dominate an integrand. Then, the integrability condition is the following.

Assumption 2.5. *(New integrability condition) Let $t < u < T$ and let J be an open interval containing u. Then there exists a stochastic bound Z with finite expectation, such that*

$$\sup_{u' \in J} \left\{ e^{-\int_t^T \left(\mathbb{1} - e_k + \left(1_{(s \le u')} + 1_{(s > u)}\right)e_k\right)^\top \Gamma(s) X(s)\,\mathrm{d}s} e_k^\top \Gamma(u') X(u') \right\} \le Z. \tag{2.10}$$

Proof of Theorem 2.3. (Outline) The theorem clearly holds for $u = t$. We prove the case $c = 0$ and $u < T$. So, assume that $t < u \le r < T$. Define now the matrix function $\tilde{\Gamma}(t,u,r)$ by

$$\tilde{\Gamma}_{ji}(t,u,r) = \begin{cases} \Gamma_{ji}(t), & j \ne k,\\ (1_{(-\infty,u]}(t) + 1_{(r,\infty)}(t))\Gamma_{ji}(t), & j = k, \end{cases}$$

for $j = 1, \ldots, p$ and $i = 1, \ldots, d$. Notice that $\Gamma(t) = \tilde{\Gamma}(t, u, r)$ when $u = r$. We consider k fixed throughout the proof and suppress the functions' dependence on k. First, using the definition of $\tilde{\Gamma}$, and then applying Theorem 2.2,

$$\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \left(\mathbb{1} - e_k + \left(1_{(s \leq u)} + 1_{(s > r)}\right) e_k\right)\Gamma(s) X(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] \\
&= \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \tilde{\Gamma}(s,u,r) X(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] \\
&= e^{\tilde{\phi}(t,T,u) + \tilde{\psi}(t,T,u)^\top X(t)}.
\end{aligned} \tag{2.11}$$

Here, $\tilde{\phi}$ and $\tilde{\psi}$ solve the differential equations (2.7) with $\tilde{\gamma} = \mathbb{1}^\top \tilde{\Gamma}$ in the place of γ, and we have added u as an argument to make clear the dependence of u in the solution. The solution also depends on r, but that is suppressed in the notation.

We apply $-\frac{\partial}{\partial u}$ on both sides of (2.11). On the left hand side we obtain,

$$\begin{aligned}
&-\frac{\partial}{\partial u} \mathrm{E}\left[e^{-\int_t^T \left(\mathbb{1} - e_k + \left(1_{(s \leq u)} + 1_{(s > r)}\right) e_k\right)^\top \Gamma(s) X(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] \\
&= -\frac{\partial}{\partial u} \mathrm{E}\left[e^{-\int_t^T \left(\mathbb{1} - e_k + 1_{(s > r)} e_k\right)^\top \Gamma(s) X(s)\,\mathrm{d}s - \int_t^u e_k^\top \Gamma(s) X(s)\,\mathrm{d}s} \,\middle|\, \mathcal{F}(t) \right] \\
&= \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \tilde{\Gamma}(s,u,r) X(s)\,\mathrm{d}s} e_k^\top \Gamma(u) X(u) \,\middle|\, \mathcal{F}(t) \right],
\end{aligned}$$

where differentiation and expectation can be interchanged by the integrability condition (2.10). When $u = r$, this is the left hand side of (2.8) with $c = 0$. For differentiation, we have implicitly assumed that $e_k^\top \Gamma(t)$ is continuous in a neighbourhood around u, however this is not a necessary assumption for the theorem.

Applying $-\frac{\partial}{\partial u}$ on the right hand side of (2.11), we obtain

$$\begin{aligned}
&-\frac{\partial}{\partial u} e^{\tilde{\phi}(t,T,u) + \tilde{\psi}(t,T,u)^\top X(t)} \\
&= e^{\tilde{\phi}(t,T,u) + \tilde{\psi}(t,T,u)^\top X(t)} \left(-\frac{\partial}{\partial u} \tilde{\phi}(t,T,u) + X(t)^\top \left(-\frac{\partial}{\partial u} \tilde{\psi}(t,T,u) \right) \right).
\end{aligned}$$

Let $\tilde{A}(t,T,u) = -\frac{\partial}{\partial u} \tilde{\phi}(t,T,u)$ and $\tilde{B}(t,T,u) = -\frac{\partial}{\partial u} \tilde{\psi}(t,T,u)$. We find, using (2.7),

$$\begin{aligned}
&\tilde{A}(t,T,u) \\
&= -\frac{\partial}{\partial u} \int_T^t \left(-\frac{1}{2} \tilde{\psi}(s,T,u)^\top a(s) \tilde{\psi}(s,T,u) - b(s)^\top \tilde{\psi}(s,T,u) \right) \mathrm{d}s \\
&= \int_T^t \left(-\tilde{\psi}(s,T,u)^\top a(s) \tilde{B}(s,T,u) - b(s)^\top \tilde{B}(s,T,u) \right) \mathrm{d}s.
\end{aligned}$$

Then, for $i = 1, \dots, d$,

$$
\begin{aligned}
&\tilde{B}_i(t,T,u) \\
&= -\frac{\partial}{\partial u}\int_T^t \left(-\frac{1}{2}\tilde{\psi}(s,T,u)^\top \alpha_i(s)\tilde{\psi}(s,T,u) - \beta_i(s)^\top \tilde{\psi}(s,T,u) + \tilde{\gamma}_i(s,u,r)\right) \mathrm{d}s \\
&= \int_T^t \left(-\tilde{\psi}(s,T,u)^\top \alpha_i(s)\tilde{B}(s,T,u) - \beta_i(s)^\top \tilde{B}(s,T,u)\right) \mathrm{d}s - \frac{\partial}{\partial u}\int_T^t \sum_{j=1}^p \tilde{\Gamma}_{ji}(s,u,r)\,\mathrm{d}s \\
&= \int_T^t \left(-\tilde{\psi}(s,T,u)^\top \alpha_i(s)\tilde{B}(s,T,u) - \beta_i(s)^\top \tilde{B}(s,T,u)\right) \mathrm{d}s - 1_{(t<u)}\frac{\partial}{\partial u}\int_u^t \Gamma_{ki}(s)\,\mathrm{d}s \\
&= \int_T^t \left(-\tilde{\psi}(s,T,u)^\top \alpha_i(s)\tilde{B}(s,T,u) - \beta_i(s)^\top \tilde{B}(s,T,u)\right) \mathrm{d}s + 1_{(t<u)}\Gamma_{ki}(u).
\end{aligned}
$$

In particular, $\tilde{A}(T,T,u) = 0$ and $\tilde{B}(T,T,u) = 0$, and since the integrands are linear in $\tilde{B}$ for $t > u$, we have $\tilde{A}(t,T,u) = 0$ and $\tilde{B}(t,T,u) = 0$ for all $t > u$ as well.

Now let $r = u$. Then $\Gamma = \tilde{\Gamma}$ and thus $\psi = \tilde{\psi}$ and $\phi = \tilde{\phi}$, where ϕ and ψ are solutions to (2.7). Letting $A^k(t,T,u) = \tilde{A}(t,T,u)$ and $B^k(t,T,u) = \tilde{B}(t,T,u)$, the system of differential equations (2.9) is obtained, when $c = 0$.

For $u = T$, the result holds as well, which can be shown by taking limits as $u \nearrow T$. The left and right hand side of (2.8) are continuous in u when $u < T$, and it can be shown that the continuity also holds in the limit as $u \nearrow T$.

For a general integrable function c, we can rewrite in terms of the functions ϕ^0, ψ, $A^{k,0}$ and B^k corresponding to the case $c = 0$. (The functions ψ and B^k do not depend on c.) We get,

$$
\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top (c(s)+\Gamma(s)X(s))\,\mathrm{d}s} e_k^\top\left(c(u) + \Gamma(u)X(u)\right)\Big|\,\mathcal{F}(t)\right] \\
&= e^{-\int_t^T \mathbb{1}^\top c(s)\,\mathrm{d}s}\Big(\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \Gamma(s)X(s)\,\mathrm{d}s}\Big|\,\mathcal{F}(t)\right] e_k^\top c(u) \\
&\qquad + \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \Gamma(s)X(s)\,\mathrm{d}s} e_k^\top \Gamma(u)X(u)\Big|\,\mathcal{F}(t)\right]\Big) \\
&= e^{-\int_t^T \mathbb{1}^\top c(s)\,\mathrm{d}s + \phi^{c=0}(t,T) + \psi(t,T)^\top X(t)}\left(e_k^\top c(u) + A^{k,c=0}(t,T,u) + B^k(t,T,u)^\top X(t)\right) \\
&= e^{\phi(t,T)+\psi(t,T)^\top X(t)}\left(A^k(t,T,u) + B^k(t,T,u)^\top X(t)\right),
\end{aligned}
$$

where the last equality sign holds whenever ϕ and A^k solve the differential equations (2.7) and (2.9), respectively. This completes the proof. □

The interesting result is the identity (2.8), and this can immediately be extended, which is done in the following corollary. Since the expectation operator is linear, we have

$$
\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s}\left(Y_k(u) + Y_l(v)\right)\Big|\,\mathcal{F}(t)\right] \\
&= \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u)\Big|\,\mathcal{F}(t)\right] + \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_l(v)\Big|\,\mathcal{F}(t)\right],
\end{aligned}
$$

for $l \in \{1, \dots, p\}$ and $v \in [t, T]$. By Theorem 2.3, the two expectations on the right hand side can be calculated, thus enabling us to find the left hand side. However, using that the system of differential equations for (A, B) is linear, we can actually calculate the left hand side directly. This is stated in a general way in the following corollary to Theorem 2.3. For the corollary, note that for a finite linear combination of elements of the type $Y_k(u)$, there exist $q \geq 1$, vectors $\kappa^1, \dots, \kappa^q \in \mathbb{R}^p$ and time points $u_1, \dots, u_q \in [t, T]$ such that the linear combination can be written as $\sum_{l=1}^{q} {\kappa^l}^\top Y(u_l)$.

Corollary 2.6. *For $q \geq 1$ let $\kappa^1, \dots, \kappa^q \in \mathbb{R}^p$ be vectors and let $u_1, \dots, u_q \in [t, T]$ be time points. If the conditions of Theorem 2.3 are satisfied for all combinations of time points $u_1, \dots, u_q$ and dimensions $k = 1, \dots, p$, then*

$$\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \sum_{l=1}^{q} {\kappa^l}^\top Y(u_l) \middle| \mathcal{F}(t) \right] \\
&= e^{\phi(t,T) + \psi(t,T)^\top X(t)} \Big(A(t,T) + B(t,T)^\top X(t) \Big),
\end{aligned}$$

where (ϕ, ψ) is a solution to (2.7) and (A, B) solves the linear system of differential equations with jumps,

$$\begin{aligned}
\frac{\partial}{\partial t} A(t,T) &= -\psi(t,T)^\top a(t) B(t,T) - b(t)^\top B(t,T), \\
A(u_l-, T) &= A(u_l, T) + {\kappa^l}^\top c(u_l), \quad l = 1, \dots, q \\
A(T,T) &= 0, \\
\frac{\partial}{\partial t} B_i(t,u) &= -\psi(t,T)^\top \alpha_i(t) B(t,u) - \beta_i(t)^\top B(t,u), \quad i = 1, \dots, d \\
B(u_l-, T) &= B(u_l, T) + {\kappa^l}^\top \Gamma(u_l), \quad l = 1, \dots, q \\
B(T,T) &= 0.
\end{aligned} \tag{2.12}$$

Proof. By linearity of the expectation operator and Theorem 2.3

$$\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \sum_{l=1}^{q} {\kappa^l}^\top Y(u_l) \middle| \mathcal{F}(t) \right] \\
&= \sum_{l=1}^{q} \sum_{k=1}^{p} \kappa_k^l \, \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) \middle| \mathcal{F}(t) \right] \\
&= e^{\phi(t,T) + \psi(t,T)^\top X(t)} \sum_{l=1}^{q} \sum_{k=1}^{p} \kappa_k^l \Big(A^k(t,T,u_l) + B^k(t,T,u_l)^\top X(t) \Big),
\end{aligned}$$

where we used Theorem 2.3 for the last equality, such that A^k and B^k solve (2.9). Now, let

$$A(t,T) = \sum_{l=1}^{q}\sum_{k=1}^{p} \kappa_k^l A^k(t,T,u_l),$$
$$B(t,T) = \sum_{l=1}^{q}\sum_{k=1}^{p} \kappa_k^l B^k(t,T,u_l),$$

and recall that $A^k(t,T,u) = 0$ and $B^k(t,T,u) = 0$ for $t > u$ and any $k = 1, \dots, d$. Then for $t \le T$,

$$\begin{aligned} A(t,T) &= \sum_{l=1}^{q} 1_{(t\le u_l)} \sum_{k=1}^{p} \Big\{ \kappa_k^l e_k^\top c(u_l) \\ &\quad + \int_{u_l}^{t} \Big(-\psi(s,T)^\top a(s)\kappa_k^l B^k(s,T,u_l) - b(s)^\top \kappa_k^l B^k(s,T,u_l) \Big)\, \mathrm{d}s \Big\} \\ &= \sum_{l=1}^{q} 1_{(t\le u_l)} {\kappa^l}^\top c(u_l) + \int_T^t \Big(-\psi(s,T)^\top a(s) B(s,T) - b(s)^\top B(s,T) \Big)\, \mathrm{d}s. \end{aligned}$$

The calculation for $B(t,T)$ is analogous, and the result is obtained. □

The functions A and B can be compared to A^k and B^k from Theorem 2.3. As is seen from the proof, they are a linear function of functions $A^k(t,T,u_l)$ and $B^k(t,T,u_l)$ that solves the linear differential equation system (2.9). Thus the linear differential equation system for A and B is also as in (2.9), except the different boundary conditions. In the system of differential equations for A and B extra jumps occur, which we can consider as gluing boundary conditions. These are exactly the boundary conditions of each of the system of differential equations for the functions A^k and B^k that add up to A and B.

The presentation in the corollary of the transformation yields another insight. We consider an affine transformation $\mathbf{Y}$ of $\mathbf{X}$, but the results also hold for the more general

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \kappa^\top X(u) \,\middle|\, \mathcal{F}(t) \right].$$

In this case, the boundary conditions in (2.12) are changed, such that $c(t) = 0$ and $\Gamma(u) = 1$. To give an intuition for the validity of this, consider for simplicity the case where $c(t) = 0$, and where $\Gamma(u)$ has a left inverse, $\Gamma^{-1}(u)$. Then apply the corollary for $q = 1$ and $\kappa = \tilde{\kappa}\Gamma^{-1}(u)$ for any $\tilde{\kappa}$. We get,

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \kappa^\top Y(u) \,\middle|\, \mathcal{F}(t) \right] = \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \tilde{\kappa}^\top X(u) \,\middle|\, \mathcal{F}(t) \right].$$

In other words, the affine transformation of $\mathbf{X}$ in the exponentiated integral, does not need to be the same as the affine transformation outside the exponentiation.

We now present a result which, to the author's knowledge, is new. Theorem 2.2 is the essential result about affine processes in a multidimensional framework. This result was used to obtain Theorem 2.3, by differentiation in a specific way. This approach can be applied again, and the following theorem is obtained.

Theorem 2.7. *Let $k, l \in \{1, \dots, p\}$ and $u, v \in [t, T]$. Assuming sufficient integrability, then*

$$
\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) Y_l(v) \middle| \mathcal{F}(t) \right] \\
&= e^{\phi(t,T) + \psi(t,T)^\top X(t)} \\
&\quad \times \Big\{ \Big(A^k(t,T,u) + B^k(t,T,u)^\top X(t) \Big) \Big(A^l(t,T,v) + B^l(t,T,v)^\top X(t) \Big) \\
&\qquad + C^{kl}(t,T,u,v) + D^{kl}(t,T,u,v)^\top X(t) \Big\},
\end{aligned}
\tag{2.13}
$$

where $\big(A^k(t,T,u), B^k(t,T,u)\big)$ and $\big(A^l(t,T,v), B^l(t,T,v)\big)$ are given by Theorem 2.3. The functions $C^{kl}(t,T,u,v)$ and $D^{kl}(t,T,u,v)$ are solutions to the following system of differential equations,

$$
\begin{aligned}
\frac{\partial}{\partial t} C^{kl}(t,T,u,v) &= -B^k(t,T,u)^\top a(t) B^l(t,T,v) - \psi(t,T)^\top a(t) D^{kl}(t,T,u,v) \\
&\quad - b(t)^\top D^{kl}(t,T,u,v), \\
C^{kl}(u \wedge v, T, u, v) &= 0, \\
\frac{\partial}{\partial t} D_i^{kl}(t,T,u,v) &= -B^k(t,T,u)^\top \alpha_i(t) B^l(t,T,v) - \psi(t,T)^\top \alpha_i(t) D^{kl}(t,T,u,v) \\
&\quad - \beta_i(t)^\top D^{kl}(t,T,u,v), \quad i = 1, \dots, n \\
D^{kl}(u \wedge v, T, u, v) &= 0,
\end{aligned}
$$

using the notation $x \wedge y = \min\{x, y\}$.

Proof. (Outline) The proof is analogous to the proof of Theorem 2.3. If either $u = t$ or $v = t$, the result follows from Theorem 2.3. We first prove the result for the case $c(t) = 0$ and $u, v < T$, so assume that $u, v \in (t, T)$ and $u \neq v$. As in the proof of Theorem 2.3, define now the matrix function $\tilde{\Gamma}^k(t, u, r)$ by

$$
\tilde{\Gamma}^k_{ji}(t,u,r) = \begin{cases} \Gamma_{ji}(t) & j \neq k \\ \big(1_{(-\infty,u]}(t) + 1_{(r,\infty)}(t)\big)\, \Gamma_{ji}(t) & j = k \end{cases},
$$

for $j = 1, \dots, p$ and $i = 1, \dots, d$. Notice that $\Gamma(t) = \tilde{\Gamma}^k(t, u, r)$ when $u = r$. An application of Theorem 2.3 yields

$$
\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \tilde{\Gamma}^k(s,u,r) X(s)\,\mathrm{d}s} e_l^\top \tilde{\Gamma}^k(v,u,r) X(v) \middle| \mathcal{F}(t) \right] \\
&= e^{\tilde{\phi}(t,T) + \tilde{\psi}(t,T)^\top X(t)} \left(\tilde{A}^l(t,T,v) + \tilde{B}^l(t,T,v)^\top X(t) \right).
\end{aligned}
$$

Here, $\tilde{\phi}$ and $\tilde{\psi}$ solve the differential equations (2.7) with $\tilde{\gamma} = \mathbb{1}^\top \tilde{\Gamma}^k$ in place of γ. The functions $\tilde{A}^l$ and $\tilde{B}^l$ solve the differential equations (2.9) with boundary conditions $\tilde{A}^l(v,T,v) = 0$ and $\tilde{B}^l(v,T,v) = e_l^\top \tilde{\Gamma}^k(v,u,r)$. Note that the functions $\tilde{\phi}$, $\tilde{\psi}$, $\tilde{A}^l$ and $\tilde{B}^l$ all depend on u. Applying $-\frac{\partial}{\partial u}$ on both sides, we obtain, after a few calculations similar to those in the proof of Theorem 2.3,

$$
\begin{aligned}
&\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \tilde{\Gamma}^k(s,u,r)X(s)\,\mathrm{d}s} e_k^\top \Gamma(u)X(u) e_l^\top \tilde{\Gamma}^k(v,u,r)X(v) \,\middle|\, \mathcal{F}(t)\right] \\
&= e^{\tilde{\phi}(t,T)+\tilde{\psi}(t,T)^\top X(t)} \\
&\quad \times \left\{ \left(-\frac{\partial}{\partial u}\tilde{\phi}(t,T) - X(t)^\top \frac{\partial}{\partial u}\tilde{\psi}(t,T)\right)\left(\tilde{A}^l(t,T,v) + \tilde{B}^l(t,T,v)^\top X(t)\right) \right. \\
&\qquad \left. - \frac{\partial}{\partial u}\tilde{A}^l(t,T,v) - X(t)^\top \frac{\partial}{\partial u}\tilde{B}^l(t,T,v) \right\}.
\end{aligned}
$$

From the proof of Theorem 2.3 it follows that $\tilde{A}^k(t,T,u) = -\frac{\partial}{\partial u}\tilde{\phi}(t,T)$ and $\tilde{B}^k(t,T,u) = -\frac{\partial}{\partial u}\tilde{\psi}(t,T)$ are solutions to the differential equations (2.9), with boundary conditions $\tilde{A}^k(u,T,u) = 0$ and $\tilde{B}^k(u,T,u) = e_k^\top \Gamma(u)$, and $\tilde{\phi}$ and $\tilde{\psi}$ in the place of ϕ and ψ.

Now, let $\tilde{C}(t,T,u,v) = -\frac{\partial}{\partial u}\tilde{A}^l(t,T,v)$ and $\tilde{D}(t,T,u,v) = -\frac{\partial}{\partial u}\tilde{B}^l(t,T,v)$. By straightforward differentiation, we find

$$
\begin{aligned}
\tilde{C}(t,T,u,v) &= -\frac{\partial}{\partial u}\int_v^t \left(-\tilde{\psi}(s,T)^\top a(s)\tilde{B}^l(s,T,v) - b(s)^\top \tilde{B}^l(s,T,v)\right) \mathrm{d}s \\
&= \int_v^t \left\{ -\tilde{B}^k(s,T,u)^\top a(s)\tilde{B}^l(s,T,v) \right. \\
&\qquad \left. - \tilde{\psi}(s,T)^\top a(s)\tilde{D}(t,T,u,v) - b(s)^\top \tilde{D}(s,T,u,v) \right\} \mathrm{d}s.
\end{aligned}
$$

For $\tilde{D}$ we find $\tilde{D}_i$ for $i = 1, \ldots, d$,

$$
\begin{aligned}
\tilde{D}_i(t,T,u,v) &= -\frac{\partial}{\partial u}\int_v^t \left(-\tilde{\psi}(s,T)^\top \alpha_i(s)\tilde{B}^l(s,T,v) - \beta_i(s)^\top \tilde{B}^l(s,T,v)\right) \mathrm{d}s \\
&= \int_v^t \left\{ -\tilde{B}^k(s,T,u)^\top \alpha_i(s)\tilde{B}^l(s,T,v) \right. \\
&\qquad \left. - \tilde{\psi}(s,T)^\top \alpha_i(s)\tilde{D}(s,T,u,v) - \beta_i(s)^\top \tilde{D}(s,T,u,v) \right\} \mathrm{d}s.
\end{aligned}
$$

We have $\tilde{D}(t,T,u,v) = 0$ for $t > v$. Since $\tilde{B}^k(t,T,u) = 0$ for $t > u$, we also have $\tilde{D}(t,T,u,v) = 0$ for $t > u$. Similarly $\tilde{C}(t,T,u,v) = 0$ for $t > u \wedge v$.

Let $r = u$. Then $\Gamma = \tilde{\Gamma}^k$ and thus $\phi = \tilde{\phi}$ and $\psi = \tilde{\psi}$, where (ϕ,ψ) solves (2.7). In this case, $A^i(t,T,\eta) = \tilde{A}^i(t,T,\eta)$ and $B^i(t,T,\eta) = \tilde{B}^i(t,T,\eta)$ for $(i,\eta) \in \{(k,u),(l,v)\}$, where (A^i,B^i) solves (2.9). Also, in this case, let $C^{kl} = \tilde{C}$ and $D^{kl} = \tilde{D}$. The result is now obtained for $c = 0$ and $u, v < T$.

If $u = T$, $v = T$ or $u = v$, then take limits for $u \nearrow T$, $v \nearrow T$ or $u \to v$, respectively, which yields the result, since both the left hand and right hand side of (2.13) can be shown to be continuous in the arguments u and v on $[t, T]$.

The extension to $c \neq 0$ can be done analogously to the proof of Theorem 2.3. First use the linearity of the expectation operator, second apply Theorems 2.2 and 2.3, and last verify the differential equations. This completes the proof. $\square$

2.3 Generalisations of the transformations

Theorem 2.2 provides the relation (2.6), repeated here,

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} \middle| \mathcal{F}(t) \right] = e^{\phi(t,T)+\psi(t,T)^\top X(t)},$$

where the functions ϕ and ψ can be found by solving a set of Riccati differential equations, (2.7). By a differentiation argument, as carried out in the proof, Theorem 2.3 gives us the relation (2.8), repeated here,

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) \middle| \mathcal{F}(t) \right] = e^{\phi(t,T)+\psi(t,T)^\top X(t)} \left(A^k(t,T,u) + B^k(t,T,u)^\top X(t) \right).$$

The functions A^k and B^k are given by a set of differential equations, (2.9). Since they arise through a differentiation argument, we essentially think of them as ϕ and ψ differentiated, respectively. By an application of the exact same differentiation technique, but to relation (2.8) instead of (2.6), we then obtained Theorem 2.7, which is the relation

$$\begin{aligned} \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) Y_l(v) \middle| \mathcal{F}(t) \right] &= e^{\phi(t,T)+\psi(t,T)^\top X(t)} \\ \times \Big\{ \left(A^k(t,T,u) + B^k(t,T,u)^\top X(t) \right) & \left(A^l(t,T,v) + B^l(t,T,v)^\top X(t) \right) \\ + C^{kl}(t,T,u,v) + D^{kl}(t,T,u,v)^\top X(t) \Big\}, & \end{aligned}$$

i.e. (2.13). The functions C^{kl} and D^{kl} are given by a set of differential equations. Again, they arise because of a differentiation, and we essentially think of them as A^k and B^k (or A^l and B^l) differentiated, respectively. This can also be seen from the proofs given.

There is no particular reason to stop here. We can apply the differentiation technique to (2.13), and obtain an expression for

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u) Y_l(v) Y_r(w) \middle| \mathcal{F}(t) \right],$$

for some $r \in \{1, \ldots, p\}$ and $w \in [t, T]$. To find the expression, one must apply the differentiation technique to the right hand side of (2.13). The result is obtainable, but the notation and number of differential equations grow with every differentiation, and thus becomes even more cumbersome. In principle, one can reapply the technique, and obtain expressions for any expectation of the form,

$$\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_{k_1}(u_1)\cdots Y_{k_q}(u_q) \middle| \mathcal{F}(t)\right], \tag{2.14}$$

for $k_1,\ldots,k_q \in \{1,\ldots,p\}$, $u_1,\ldots,u_q \in [t,T]$ and $q \geq 0$. We can count the number of differential equations that need to be solved when using this approach. For the expression (2.6), corresponding to the case $q = 0$, the functions ϕ and ψ must be found, which is a system of differential equations of dimension $d+1$. For the second expression, (2.8) (corresponding to $q = 1$), the functions A^k and B^k must also be found, which is $d+1$ extra dimensions, in total $2(d+1)$. For the third expression, (2.13) (corresponding to $q = 2$), A^l and B^l as well as C^{kl} and D^{kl} must be found, which is $2(d+1)$ extra equations, in total $4(d+1)$. It seems that the dimension of the system of differential equations that needs to be solved is increasing exponentially.

In the relation (2.13), one can choose $k = l$ and $u = v$, to obtain

$$\begin{aligned}\mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top Y(s)\,\mathrm{d}s} Y_k(u)^2 \middle| \mathcal{F}(t)\right] &= e^{\phi(t,T)+\psi(t,T)^\top X(t)} \\ &\times \left\{ \left(A^k(t,T,u) + B^k(t,T,u)^\top X(t)\right)^2 + C^{kk}(t,T,u,u) + D^{kk}(t,T,u,u)^\top X(t)\right\}.\end{aligned}$$

Similarly, for particular choices of $k_1,\ldots,k_q$ and $u_1,\ldots,u_q$ in (2.14), one can obtain all moments and combinations of moments of different $Y_k(u)$. Using linearity of the expectation operator, one can use this to construct any polynomial in $Y_k(u)$.

A special case of transformations are moments of affine processes. For a process $\mathbf{Y}$ consider the modified process,

$$\hat{Y}(t) = 1_{\{u,v\}}(t) Y(t) = \hat{c}(t) + \hat{\Gamma}(t) X(t),$$

where $\hat{c}(t) = 1_{\{u,v\}}(t)c(t)$ and $\hat{\Gamma}(t) = 1_{\{u,v\}}(t)\Gamma(t)$. Then it holds that

$$\mathrm{E}\left[Y_k(u) \middle| \mathcal{F}(t)\right] = \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \hat{Y}(s)\,\mathrm{d}s} \hat{Y}_k(u) \middle| \mathcal{F}(t)\right] = A^k(t,T,u) + B^k(t,T,u)^\top X(t),$$

where $\phi(t,T) = \psi(t,T) = 0$, because $\hat{c}$ and $\hat{\Gamma}$ equal zero almost surely with respect to the Lebesgue measure. The system of differential equations for A^k and B^k simplifies to

$$\begin{aligned} \frac{\partial}{\partial t} A^k(t,T,u) &= -b(t)^\top B^k(t,T,u), & A^k(u,T,u) &= e_k^\top c(u), \\ \frac{\partial}{\partial t} B^k(t,T,u) &= -\mathcal{B}(t)^\top B^k(t,T,u), & B^k(u,T,u) &= e_k^\top \Gamma(u). \end{aligned}$$

Considering the second moment, we can obtain an interpretation of the functions C^{kl} and D^{kl} as a covariance. See that

$$\begin{aligned}\mathrm{E}\left[Y_k(u)Y_l(v)\middle|\,\mathcal{F}(t)\right] &= \mathrm{E}\left[e^{-\int_t^T \mathbb{1}^\top \hat{Y}(s)\,\mathrm{d}s}\hat{Y}_k(u)\hat{Y}_l(v)\middle|\,\mathcal{F}(t)\right] \\ &= \left(A^k(t,T,u)+B^k(t,T,u)^\top X(t)\right)\left(A^l(t,T,v)+B^l(t,T,v)^\top X(t)\right) \\ &\quad + C^{kl}(t,T,u,v)+D^{kl}(t,T,u,v)^\top X(t).\end{aligned}$$

Combining with the result above, we conclude that

$$\mathrm{Cov}\left[Y_k(u),Y_l(v)\middle|\,\mathcal{F}(t)\right] = C^{kl}(t,T,u,v)+D^{kl}(t,T,u,v)^\top X(t).$$

We note here that the functions C^{kl} and D^{kl} of course depend on the functions $\hat{c}(t)$ and $\hat{\Gamma}(t)$ through $\psi(t,T)$, which is very special in this case because it is equal to zero. In general the functions C^{kl} and D^{kl} do not give the covariance of the stochastic variables $Y_k(u)$ and $Y_l(v)$.

3 Doubly stochastic decrement Markov chains

In this section, we consider so-called decrement Markov chains in finite state spaces with affine and dependent transition rates. The theorems presented above, and their generalisations, allow us to calculate transition probabilities for such doubly stochastic Markov chains. We use the notion decrement (or multiple decrement) Markov chain for the case where, for each state i, the Markov chain cannot return to state i after leaving it. This is also sometimes referred to in the literature as a hierarchical Markov chain model. The reason for the restriction to this class is, that in the case of deterministic rates, the transition probabilities can be written as integral expressions, which is necessary for the approach presented here.

Let a finite state space J be given. We associate a set of non-negative transition rates (μ_{ij}), $i,j \in \mathcal{J}$, with some of the transition rates identical to zero, such that it is a multiple decrement Markov chain. Let the set of non-zero transition rates be modelled as an affine transformation of a d-dimensional affine process $\mathbf{X}$. That is, assuming that there are p non-zero transition rates, let functions $c : \mathbb{R}_+ \to \mathbb{R}^p$ and $\Gamma : \mathbb{R}_+ \to \mathbb{R}^{p\times d}$ be given, and define $Y(t) = c(t) + \Gamma(t)X(t)$. Then each of the stochastic transition rates is modelled as an element in $\mathbf{Y}$, i.e. for each non-zero transition rate μ_{ij}, there is a dimension in $\mathbf{Y}$, k say, such that $\mu_{ij}(t) = Y_k(t)$.

Define now the stochastic process $\mathbf{Z} = (Z(t))_{t\in\mathbb{R}_+}$ with $Z(0) = 0$, and let $\mathbf{Z}$ be a Markov chain in $\mathcal{J}$ with distribution given by the transition rates (μ_{ij}), conditional on $\mathbf{X}$. That is, we have defined $\mathbf{Z}$ through the conditional distribution, given the stochastic transition rates. With $(N_{ij}(t))_{t\in\mathbb{R}_+}$, $i,j \in \mathcal{J}$, being the process that counts the number of jumps for $\mathbf{Z}$ from state i to j, the compensated process

$$N_{ij}(t) - \int_0^t 1_{(Z(s-)=i)}\mu_{ij}(s)\,\mathrm{d}s \tag{3.1}$$

is a martingale, conditional on $\mathbf{X}$.

Let the filtrations $\mathbb{F}^{\mathbf{Z}} = (\mathcal{F}^{\mathbf{Z}}(t))_{t\in\mathbb{R}_+}$ and $\mathbb{F}^{\mathbf{X}} = (\mathcal{F}^{\mathbf{X}}(t))_{t\in\mathbb{R}_+}$ be the ones generated by the processes $\mathbf{Z}$ and $\mathbf{X}$, respectively, satisfying the usual hypotheses. Let the general filtration be given by $\mathbb{F} = (\mathcal{F}(t))_{t\in\mathbb{R}_+} = \left(\mathcal{F}^{\mathbf{Z}}(t) \vee \mathcal{F}^{\mathbf{X}}(t)\right)_{t\in\mathbb{R}_+}$.

The transition probability from state i at time s to state j at time t is now $\mathcal{F}^{\mathbf{X}}(s)$-measurable, and using the tower property and the Markov property of $\mathbf{X}$, it can be written as

$$\begin{aligned} p_{ij}(s,t) &:= P\left(Z(t)=j \,|\, Z(s)=i, \mathcal{F}^{\mathbf{X}}(s)\right) \\ &= \mathrm{E}\left[P\left(Z(t)=j \,|\, Z(s)=i, \mathcal{F}^{\mathbf{X}}(\infty)\right) \middle| \mathcal{F}^{\mathbf{X}}(s)\right] \\ &= \mathrm{E}\left[p^{\mathbf{X}}_{ij}(s,t) \middle| X(s)\right], \end{aligned}$$

where $p^{\mathbf{X}}_{ij}(s,t)$ is the transition probability in the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$. From this calculation it is seen, that if we know the conditional transition probabilities, corresponding to the case of known transition rates, we can find the unconditional ones by applying the expectation operator. For hierarchical models, this is exactly the expectations appearing in Theorems 2.2, 2.3 and 2.7, and the generalisations (2.14). We can, in principle, find transition probabilities in any doubly stochastic decrement Markov chain with affine transition rates. The conditional transition probabilities $p^{\mathbf{X}}_{ij}(s,t)$ are known explicitly and can be written in sums and integrals of expressions of the type (2.14). This can be verified using e.g. Kolmogorov's backward differential equation.

In order to illustrate the method in practice, we present an example in a simple state space and show how to find the transition probabilities, by applying Theorems 2.2, 2.3 and 2.7.

Example 3.1. Let $\mathcal{J} = \{0,1,2\}$, and assume that the non-zero transition rates are μ_{01}, μ_{02} and μ_{12}. This could be a life insurance disability model with state 0, 1 and 2 corresponding to *active*, *disabled* and *dead*, as shown in Figure 1. The transition rates are modelled by $(\mu_{01}(t), \mu_{02}(t), \mu_{12}(t))^\top = Y(t) = c(t) + \Gamma(t)X(t)$, for some affine process $\mathbf{X}$ and functions c and Γ.

Conditional on $\mathbf{X}$, the transition probabilities are known explicitly, and for state 0 we

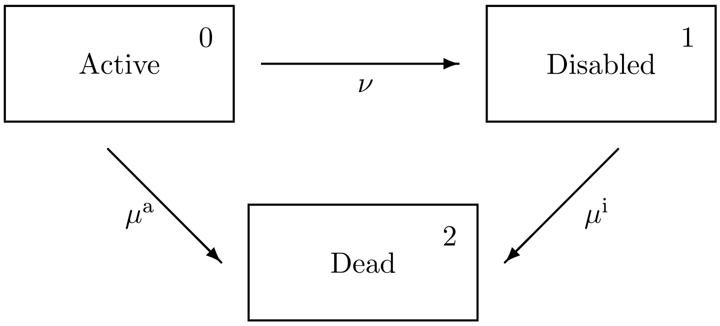

Figure 1: *State space for the disability model.*

have

$$
\begin{aligned}
p_{00}^{\mathbf{X}}(t,s) &= e^{-\int_t^s (\mu_{01}(\tau)+\mu_{02}(\tau))\,\mathrm{d}\tau}, \\
p_{01}^{\mathbf{X}}(t,s) &= \int_t^s e^{-\int_t^u (\mu_{01}(\tau)+\mu_{02}(\tau))\,\mathrm{d}\tau} \mu_{01}(u) e^{-\int_u^s \mu_{12}(\tau)\,\mathrm{d}\tau}\,\mathrm{d}u, \\
p_{02}^{\mathbf{X}}(t,s) &= \int_t^s e^{-\int_t^u (\mu_{01}(\tau)+\mu_{02}(\tau))\,\mathrm{d}\tau} \left(\mu_{02}(u) + \mu_{01}(u) \int_u^s e^{-\int_u^v \mu_{12}(\tau)\,\mathrm{d}\tau} \mu_{12}(v)\,\mathrm{d}v \right) \mathrm{d}u \\
&= \int_t^s p_{00}^{\mathbf{X}}(t,u)\mu_{02}(u)\,\mathrm{d}u + \int_t^s p_{01}^{\mathbf{X}}(t,u)\mu_{12}(u)\,\mathrm{d}u,
\end{aligned}
\tag{3.2}
$$

and they can be verified by e.g. Kolmogorov's differential equations. Define

$$
I_{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I_{(2,u)}(t) = \begin{bmatrix} 1_{(t\leq u)} & 0 & 0 \\ 0 & 1_{(t\leq u)} & 0 \\ 0 & 0 & 1_{(t>u)} \end{bmatrix},
$$

and let the processes $\mathbf{Y}_{(1)}$ and $\mathbf{Y}_{(2,u)}$ be given by

$$
\begin{aligned}
Y_{(1)}(t) &= I_{(1)} Y(t) = c_{(1)}(t) + \Gamma_{(1)}(t) X(t), \\
Y_{(2,u)}(t) &= I_{(2,u)}(t) Y(t) = c_{(2,u)}(t) + \Gamma_{(2,u)}(t) X(t),
\end{aligned}
$$

where $c_{(1)} = I_{(1)} c$, $\Gamma_{(1)} = I_{(1)} \Gamma$, $c_{(2,u)} = I_{(2,u)} c$ and $\Gamma_{(2,u)} = I_{(2,u)} \Gamma$. With these definitions, we have

$$
\begin{aligned}
\mathbb{1}^\top Y_{(1)}(t) &= \mu_{01}(t) + \mu_{02}(t), \\
\mathbb{1}^\top Y_{(2,u)}(t) &= 1_{(t\leq u)} \left(\mu_{01}(t) + \mu_{02}(t)\right) + 1_{(t>u)} \mu_{12}(t)
\end{aligned}
$$

and the above conditional probabilities can be rewritten,

$$\begin{aligned}
p_{00}^{\mathbf{X}}(t,s) &= e^{-\int_t^s \mathbb{1}^\top Y_{(1)}(\tau)\,\mathrm{d}\tau}, \\
p_{01}^{\mathbf{X}}(t,s) &= \int_t^s e^{-\int_t^s \mathbb{1}^\top Y_{(2,u)}(\tau)\,\mathrm{d}\tau} Y_{(2,u),1}(u)\,\mathrm{d}u, \\
p_{02}^{\mathbf{X}}(t,s) &= \int_t^s e^{-\int_t^u \mathbb{1}^\top Y_{(1)}(\tau)\,\mathrm{d}\tau} Y_{(1),2}(u)\,\mathrm{d}u \\
&\quad + \int_t^s \int_t^v e^{-\int_t^v \mathbb{1}^\top Y_{(2,u)}(\tau)\,\mathrm{d}\tau} Y_{(2,u),1}(u) Y_{(2,u),3}(v)\,\mathrm{d}u\,\mathrm{d}v
\end{aligned} \tag{3.3}$$

where we use the notation $Y_{(2,u),i}(t)$ for the ith entry in the vector $Y_{(2,u)}(t)$. The real transition probabilities can then be found as the conditional expectations,

$$p_{ij}(t,s) = \mathrm{E}\left[p_{ij}^{\mathbf{X}}(t,s) \middle| X(t) \right]. \tag{3.4}$$

Insertion of (3.3) into (3.4) and using linearity and interchanging expectation and integration, we find

$$p_{00}(t,s) = e^{\phi_{(1)}(t,s)+\psi_{(1)}(t,s)^\top X(t)}, \tag{3.5}$$

$$p_{01}(t,s) = \int_t^s e^{\phi_{(2,u)}(t,s)+\psi_{(2,u)}(t,s)^\top X(t)} \left(A_{(2,u)}^1(t,s,u) + B_{(2,u)}^1(t,s,u)^\top X(t) \right) \mathrm{d}u, \tag{3.6}$$

$$\begin{aligned}
p_{02}(t,s) &= \int_t^s e^{\phi_{(1)}(t,s)+\psi_{(1)}(t,s)^\top X(t)} \left(A_{(1)}^2(t,u,u) + B_{(1)}^2(t,u,u)^\top X(t) \right) \mathrm{d}u \\
&\quad + \int_t^s \int_t^v e^{\phi_{(2,u)}(t,v)+\psi_{(2,u)}(t,v)^\top X(t)} \Bigg(\left(A_{(2,u)}^1(t,v,u) + B_{(2,u)}^1(t,v,u)^\top X(t) \right) \\
&\qquad \times \left(A_{(2,u)}^3(t,v,v) + B_{(2,u)}^3(t,v,v)^\top X(t) \right) \\
&\qquad + C_{(2,u)}^{1,3}(t,v,u,v) + D_{(2,u)}^{1,3}(t,v,u,v)^\top X(t) \Bigg) \mathrm{d}u\,\mathrm{d}v.
\end{aligned} \tag{3.7}$$

We used Theorem 2.2 for (3.5), Theorem 2.3 for (3.6) and the first part of (3.7), and Theorem 2.7 for the second part of (3.7). These theorems provide us with differential equations for the functions ϕ, ψ, A, B, C and D. In order to calculate the integrals numerically, the integrand in (3.6) and the integrand in the first part of (3.7) must be calculated for each $u \in [t,s]$. Likewise, the integrand in the second part of (3.7) must be calculated for each $u, v \in [t,s]$ where $u \leq v$. In practice the expressions are found for u and v in a discretized grid, and these can then be used to carry out the numerical integration. ○

3.1 Numerical efficiency

A transition probability in a doubly stochastic Markov chain can be calculated with Monte Carlo methods, and it can also be characterised by a partial differential equation.

In this section we extend Example 3.1 with an actual model, and calculate the transition probabilities with the method proposed. We discuss the advantages compared to Monte Carlo and PDE methods.

Example 3.2. (Example 3.1 continued.) Let the stochastic mortality be defined as

$$\begin{aligned}\mu_{02}(t) &= X_1(t) + X_2(t)c_1^{x+t}\\ \mathrm{d}X_i(t) &= -a_i X_i(t)\,\mathrm{d}t + \sigma_i\,\mathrm{d}W_i(t)\end{aligned}$$

for $i = 1, 2$, where x is the age at time $t = 0$. Here, the $\mathbf{X}_1$ and $\mathbf{X}_2$ are Ornstein-Uhlenbeck processes. This model is proposed in [12], in which other affine stochastic mortality models are studied as well. The mortality should be non-negative, which is not satisfied by this model. However, as is pointed out in [12], this drawback is not considered severe enough in practice to disregard the model, and we accept it. It is also pointed out that the probability of $\mathbf{X}_i$ becoming negative is negligible.

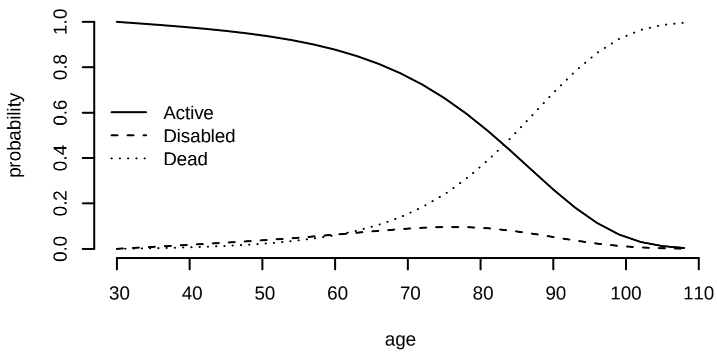

Figure 2: *Transition probabilities, starting in the active state at age 30.*

For simplicity, we choose to model $\mu_{02}(t)$ and $\mu_{12}(t)$ in a similar way, with correlation between the transition rates. Let the disability rate be given as

$$\begin{aligned}\mu_{01}(t) &= X_3(t) + X_4(t)c_2^{x+t}\\ \mathrm{d}X_i(t) &= -a_i X_i(t)\,\mathrm{d}t + \sigma_i\left(\sqrt{1-\rho_1^2}\,\mathrm{d}W_i(t) + \rho_1\,\mathrm{d}W_{i-2}(t)\right),\end{aligned}$$

for $i = 3, 4$. Thus, by choice of ρ_1 it is correlated with the mortality rate as active. Also, let the mortality rate as disabled be given as

$$\begin{aligned}\mu_{12}(t) &= X_5(t) + X_6(t)c_3^{x+t}\\ \mathrm{d}X_i(t) &= -a_i X_i(t)\,\mathrm{d}t + \sigma_i\left(\sqrt{1-\rho_2^2}\,\mathrm{d}W_i(t) + \rho_2\,\mathrm{d}W_{i-4}(t)\right),\end{aligned}$$

μ_{02}		μ_{01}		μ_{12}	
c_1	1.09144	c_2	1.09144	c_3	1.09144
a_1	0.028	a_3	0.028	a_5	0.028
a_2	0.0046	a_4	0.0046	a_6	0.0046
σ_1	$1.79 \cdot 10^{-5}$	σ_3	$1.5 \cdot 1.79 \cdot 10^{-5}$	σ_3	$10 \cdot 1.79 \cdot 10^{-5}$
σ_2	$3.83 \cdot 10^{-7}$	σ_4	$1.5 \cdot 3.83 \cdot 10^{-7}$	σ_4	$10 \cdot 3.83 \cdot 10^{-7}$
$X_1(0)$	$9.31 \cdot 10^{-5}$	$X_3(0)$	$1.5 \cdot 9.31 \cdot 10^{-5}$	$X_5(0)$	$10 \cdot 9.31 \cdot 10^{-5}$
$X_2(0)$	$2.19 \cdot 10^{-5}$	$X_4(0)$	$1.5 \cdot 2.19 \cdot 10^{-5}$	$X_6(0)$	$10 \cdot 2.19 \cdot 10^{-5}$
		ρ_1	0.5	ρ_2	0.5

Table 1: *Parameters for the stochastic transition rates. The disability rate μ_{01} is in distribution equal to $1.5\mu_{02}$ and the mortality as disabled is in distribution equal to $10\mu_{02}$. We have added a correlation. These intensities are stylised versions of what could be used in practice.*

for $i = 5, 6$. Then, if $\rho_2 \neq 0$ it is correlated with the mortality rate. In particular, if $\rho_1 \neq 0$ and $\rho_2 \neq 0$, the mortality as disabled and the disability rate are correlated. Other dependency structures could also have been chosen.

By specifying $\Gamma(t)$ as

$$\Gamma(t) = \begin{bmatrix} 1 & c_1^{x+t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & c_2^{x+t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & c_3^{x+t} \end{bmatrix},$$

we can write $(\mu_{02}(t), \mu_{01}(t), \mu_{12}(t))^\top = Y(t) = \Gamma(t)X(t)$ where $\mathbf{X}$ is the 6-dimensional affine process specified above. We choose parameters for illustrative purpose, partly taken from [12], and they are given in Table 1.

In Figure 2 the transition probabilities are seen, starting as an active 30 year old. These are found calculating (3.5) – (3.7) numerically, by first finding ϕ, ψ, A, B, C and D by numerically solving their differential equations, and afterwards carrying out a numerical integration. In this example, some of the differential equations can be solved analytically, which is well known for Ornstein-Uhlenbeck processes, see e.g. [12]. We have solved them all numerically though, in order to present an example that generally applies to affine processes.

The differential equations are solved using the 4th order Runge-Kutta method, and the integrals in (3.5) – (3.7) are solved using a 4th order Simpson integration method. The main work is done in the `C` programming language, and the rest in `R`. With the numerical methods used, the total convergence can be shown to be of order 4 in step size, and this is seen in the left part of Figure 3. The same step size is used for solving the differential equations and the integrals. From the right part of Figure 3, we see that the convergence order in time is close to 1.5. We also see that when using step size 0.9, we can calculate

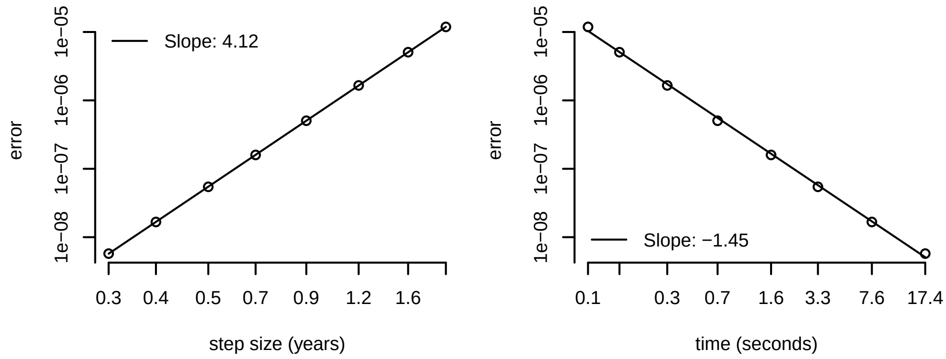

Figure 3: *Convergence of the transition probabilities of a 30 year old to age 80,* $p_{00}(0,50) + p_{01}(0,50) + p_{01}(0,50)$ *towards 1, shown with logarithmic axes. To the left we see that the convergence is of order 4 in step size. To the right the same plot is shown, but with time on the first axis instead of step size. This illustrates that the convergence in time seems to be of order 1.5.*

$p_{00}(0,50)$, $p_{01}(0,50)$ and $p_{02}(0,50)$ in 0.7 seconds, and that total error is smaller than 10^{-6}. The error and convergence order vary a bit for different ages. The calculation time increases with time-span, so it is significantly faster to calculate e.g. $p_{ij}(0,10)$ than $p_{ij}(0,50)$. ○

3.1.1 Monte Carlo and PDE methods

The stochastic process $\mathbf{X}$ can be simulated with Monte Carlo methods, and with each simulation, the transition probabilities can be calculated, e.g. by solving Kolmogorov's differential equations. Monte Carlo methods typically yields a convergence in step size of 0.5, and since one has to solve Kolmogorov's differential equations for each simulation, the convergence in time will be smaller. Thus, in our example above, exploiting the affine structure seems advantageous.

The transition probability can also be characterised by a partial differential equation: Since

$$\mathrm{E}\left[1_{(Z(t)=j)} \middle| \mathcal{F}(s)\right] = \sum_{i \in \mathcal{J}} 1_{(Z(s)=i)} p_{ij}(s,t)$$

is a martingale, Itô's formula can be applied in order to obtain a PDE that can be solved, typically by numerical methods. For more details about this approach see e.g. [3], where expectations of more general functions than the indicator function are considered. This partial differential equation has a time dimension and a dimension for each underlying process. Thus if $\mathbf{X}$ is d-dimensional, the PDE is $(d+1)$-dimensional. In Example 3.2 the process $\mathbf{X}$ is 6-dimensional, thus we obtain a 7-dimensional PDE. In practice one would use Monte Carlo methods instead of solving a 7-dimensional PDE with finite difference methods, since usually this is significantly faster. See e.g. Section 80.16 in [13], where it is stated that if there are 4 or more dimensions, a PDE problem is usually faster solved with Monte Carlo methods instead.

In Example 3.2, we considered a state space $\mathcal{J}$ with 3 states, and an affine process $\mathbf{X}$ with 6 dimensions. The PDE method suffers from the high dimension of $\mathbf{X}$, which is not a problem if the affine structure is exploited. On the other hand, if $\mathbf{X}$ is relatively simple, i.e. if it is 3-dimensional or less, the PDE method is applicable, and here a more complex state space $\mathcal{J}$ can be considered. If there are a lot of states in $\mathcal{J}$, one must solve high-dimensional integrals. Also, if the model is not hierarchical, e.g. if recovery from disability is possible, the affine methods do not work. The PDE method can handle a complex state space $\mathcal{J}$. To sum up, we can say that if the state space $\mathcal{J}$ is simple (and $\mathbf{X}$ is any affine process), the differential equations presented in this article are preferred, and on the other hand, if the stochastic process $\mathbf{X}$ is simple (and $\mathcal{J}$ is any state space), the PDE method is preferred.

4 An application in life insurance

We adopt the setup from Section 3 above, and consider a life insurance contract where $\mathbf{Z}$ describes the state of the insured. We present a brief example which shows that the methods from Example 3.1 also can be used for calculating the expected present value of a life insurance contract. This idea is further explored in [2] wherein affine processes are applied in life insurance, and in particular a surrender modelling example is studied.

Example 4.1. Let $\mathbf{Z}$ be as in Example 3.1, i.e. we consider the disability model shown in Figure 1. The interest rate process $(r(t))_{t\in\mathbb{R}_+}$ is allowed to be stochastic, and this is modelled jointly with the transition rates,

$$(r(t), \mu_{01}(t), \mu_{02}(t), \mu_{12}(t))^\top = c(t) + \Gamma(t)X(t).$$

We have thus specified a model where the interest and transition rates can have any dependent or independent affine structure. For applications, one could argue that the interest rate might not be dependent on the transition rates, however, because it does not increase the complexity of the mathematics in any significant way, the general case is considered here.

We specify a life insurance contract by the payments, given by the accumulated payment process $\mathbf{B}$, satisfying

$$\mathrm{d}B(t) = b_0(t)1_{(Z(t)=0)}\,\mathrm{d}t + b_1(t)1_{(Z(t)=1)}\,\mathrm{d}t + b_{02}(t)\,\mathrm{d}N_{02}(t) + b_{12}(t)\,\mathrm{d}N_{12}(t).$$

Here, $b_i(t)$ are continuous payments while in state i which, when negative, corresponds to premium payments. The functions $b_{ij}(t)$ are the payments upon a transition from state i to j at time t. The payment functions are deterministic, and we assume that there are no payments after time T.

Conditional on the transition rates and that the policyholder is active, we find the expected present value of the future payments as we do in the setup with deterministic interest and transition rates,

$$V^{\mathbf{X}}(t) = \int_t^T e^{-\int_t^s r(\tau)\,\mathrm{d}\tau} \Big(p_{00}^{\mathbf{X}}(t,s)\left(b_0(s) + \mu_{02}(s)b_{02}(s)\right) \\ + p_{01}^{\mathbf{X}}(t,s)\left(b_1(s) + \mu_{12}(s)b_{12}(s)\right)\Big)\,\mathrm{d}s,$$

see e.g. [2]. Here, the conditional transition probabilities $p_{0j}^{\mathbf{X}}(t,s)$ are from (3.3). The prospective reserve is now found by taking expectation, $V(t) = \mathrm{E}\left[V^{\mathbf{X}}(t)\middle|\,\mathcal{F}(t)\right]$. By insertion of $p_{0j}^{\mathbf{X}}(t,s)$ and then interchanging expectation and integration, the reserve can be calculated analogously to the way we found the transition probabilities in Examples 3.1 and 3.2. ○

Acknowledgements

I would like to thank my ph.d.-supervisors Thomas Møller and Mogens Steffensen for their guidance and support during this work.

References

[1] Enrico Biffis. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics & Economics*, 37:443–468, 2005.

[2] Kristian Buchardt. Dependent interest and transition rates in life insurance. *Insurance: Mathematics & Economics*, 55:167–179, 2014.

[3] Mikkel Dahl. Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: Mathematics & Economics*, 35 (1):113–136, 2004.

[4] Mikkel Dahl and Thomas Møller. Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: Mathematics & Economics*, 39 (2):193–217, 2006.

[5] Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68, No. 6:1343–1376, 2000.

[6] Damir Filipovic. *Term structure models.* Spring Finance. Springer, 2009.

[7] Jarrow, R. A., Lando D., and Turnbull, S. M. A Markov model for the term structure of credit risk spreads. *Journal of Finance*, 50:53–85, 1995.

[8] Holger Kraft and Mogens Steffensen. Bankruptcy, counterparty risk, and contagion. *Review of Finance*, pages 209–252, 2007.

[9] David Lando. On cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120, 1998.

[10] Thomas Møller and Mogens Steffensen. *Market-valuation methods in life and pension insurance.* International Series on Actuarial Science. Cambridge University Press, 2007.

[11] Ragnar Norberg. A theory of bonus in life insurance. *Finance and Stochastics*, 3:373–390, 1999.

[12] David F. Schrager. Affine stochastic mortality. *Insurance: Mathematics & Economics*, 38 (1):81–97, 2006.

[13] Paul Wilmott. *Paul Wilmott on quantitative finance.* John Wiley & Sons, 2nd edition, 2006.