



On the calculation of prospective and retrospective reserves in non-Markov models

Marcus C. Christiansen¹

Received: 26 February 2021 / Accepted: 16 April 2021 / Published online: 4 May 2021
© The Author(s) 2021

Abstract

Almost all life and health insurance models in the actuarial literature use either a Markov assumption or a semi-Markov assumption. This paper shows that non-Markov modelling is also feasible and presents suitable numerical and statistical tools for the calculation of prospective and retrospective reserves. A central idea is to base the calculation of reserves on forward and backward transition rates. Feasible estimators for the forward transition rates have been recently suggested in the medical statistics literature. This paper slightly extends them according to insurance needs and newly introduces symmetric estimators for backward transition rates. Only few adjustments are actually needed in the classical insurance formulas when switching from Markov modelling to as-if-Markov evaluations in order to avoid model risk.

Keywords Life and health insurance · Semi-Markov model · Kolmogorov forward equation · Landmark Aalen–Nelson estimator · Landmark Aalen–Johansen estimator

1 Introduction

Markov and semi-Markov modelling is the predominant approach in life and health insurance, even though there are numerous examples where the Markov assumptions are actually not satisfied. A central reason for that is a lack of feasible alternatives. However, forcing a Markov assumption on non-Markov data produces systematic model risk. The size of this systematic risk is hard to quantify and mostly unknown. Non-Markov modelling poses numerical and statistical challenges. Concerning the numerical issues, we show that non-Markov modelling is actually not harder than Markov modelling if actuarial reserves are calculated based on suitable forward and backward transition rates. Furthermore, we show that the statistical estimation

✉ Marcus C. Christiansen
marcus.christiansen@uni-oldenburg.de

¹ Institut für Mathematik, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany

of forward transition rates is indeed feasible by means of the so-called landmark Nelson–Aalen estimator, which was recently suggested in the medical statistics literature. We additionally develop a symmetric estimator for the backward transition rates, which are needed for the calculation of retrospective reserves. Furthermore, we slightly expand the information model according to the needs in insurance.

Parallel to Markov modelling we establish as-if-Markov modelling, which evaluates insurance liabilities conditional on the current Markov information only but without actually making the Markov assumption. We show that few adjustments to the classical Markov formulas are sufficient in order that prospective and retrospective reserves are consistently estimated even on non-Markov data.

Our approach for the numerical calculation of prospective reserves bases on a system of two forward differential equations. The first forward equation is for the forward state occupation probabilities and expands the Kolmogorov forward equation to non-Markov cases. The second forward equation is derived from an explicit expected cash flow representation of individual insurance contracts. For the numerical calculation of retrospective reserves, we develop backward equations as time-reversed forward equations. These backward equations are distinctly different from the Kolmogorov backward equation and should not be confused with it. Our time-reversion concept, which seems to be completely new in the actuarial literature, allows us to represent and calculate retrospective reserves fully symmetrically to prospective reserves.

A key step in this paper is the statistical estimation of forward and backward transition rates from observed data. The estimation of transition rates for Markov multistate models has a long tradition in the statistics literature. For nonparametric estimations, the Nelson–Aalen estimator is the preferred choice, since it can handle left-truncations and right-censoring. Recently, Putter and Spitoni [12] introduced a so-called landmark Nelson–Aalen estimator that extends the concept of the Nelson–Aalen estimator to right-censored non-Markov data. This landmark Nelson–Aalen estimator estimates a specific class of forward transition rates. The consistency proof of Putter and Spitoni [12] turned out to be incomplete, but the gap was recently closed by Overgaard [11] and Niessl et al. [8]. We rely on the results of Overgaard [10, 11] and expand the conditioning current information in the forward transition rates according to insurance needs. Moreover, we newly introduce a time-reversed landmark Nelson–Aalen estimator that estimates backward transition rates. We do not claim to comprehensively solve all statistical challenges. Our estimators primarily serve as a proof of concept, but we cover already various useful examples. We recommend to consult in addition the statistical literature on landmark estimation.

Insurance cash flows may contain both, sojourn payments for staying in a certain state and transition payments upon state changes. In case that there are no transition payments in an insurance contract, transition rates are not necessarily needed and it suffices to work with state occupation probabilities only. A popular nonparametric estimator for state occupation probabilities in Markov models is the Aalen–Johansen estimator, which is defined as the solution of the Kolmogorov forward equation with respect to the Nelson–Aalen estimator. Datta and Satten [4] observed that even for non-Markov, randomly right-censored data the Aalen–Johansen estimator still

provides consistent estimates. This fact was already used in Guibert and Planchet [7] for the calculation of prospective reserves in long-term care insurance with sojourn payments only and no transition payments.

In certain cases, our forward rates and forward equations correspond to forward concepts in Norberg [9], Buchardt [1] and Buchardt et al. [2], but aims and scope are different. The original notion of forward transition rates was to model market prices of biometric risk, whereas we focus on the real-world probability distribution. Norberg [9] and Buchardt [1] assume certain Markov structures, which we try to completely avoid. Buchardt et al. [2] define artificial forward rates that are meant for efficient numerical computations of prospective reserves. Their aim is to use the classical formulas also for non-Markov data, whereas we suggest to adjust the classical formulas.

The paper is structured as follows. In Sect. 2 we define the random pattern of states of an individual life or health insurance policy as a stochastic jump process. Section 3 discusses the information that an insurer is actually conditioning on in the calculation of prospective and retrospective reserves. Section 4 clarifies the formal definition of the stochastic differential equation notation in this paper. In Sect. 5 the insurance cash flow is mathematically defined. Section 6 presents a general extension of the Kolmogorov forward equation to non-Markov frameworks. The statistical estimation of forward and backward transition rates from data is discussed in Sect. 7. Section 8 explains the calculation of prospective and retrospective reserves based on the estimators from the Sect. 7. Section 9 concludes. An Appendix contains the proofs of all presented theorems.

2 The random pattern of states of the insured

At each point in time the insurer assigns a state to each individual insured that describes the current health status and contract status. We describe the random pattern of states of the insured by a right-continuous jump process $Z = (Z(t))_{t \geq 0}$ on a finite state space \mathcal{Z} . Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the underlying probability space. We additionally set $Z_{0-} := Z_0$. We define state indicator processes I_i and transition counting processes N_{ij} by

$$I_i(t) := \mathbb{1}_{\{Z(t)=i\}}, \quad i \in \mathcal{Z},$$

$$N_{ij}(t) := \#\{s \in (0, t] : Z(s-) = i, Z(s) = j\}, \quad i, j \in \mathcal{Z}, i \neq j,$$

which are processes with right-continuous paths. We assume that

$$\mathbb{E}[(N_{ij}(t))^2] < \infty, \quad t \geq 0, i, j \in \mathcal{Z}, i \neq j. \tag{1}$$

This condition implies in particular that the number of jumps of Z is almost surely finite on finite intervals. We can represent I_i as

$$I_i(t) = I_i(0) + \sum_{j:j \neq i} (N_{ji}(t) - N_{ij}(t)), \quad t \geq 0, i \in \mathcal{Z}. \tag{2}$$

3 The information model

We generally assume that we are currently at time $s \geq 0$. This parameter s is fixed in this paper, so we largely hide it in our notation. The available information on the state process of the insured at time s is given by the sigma-algebra

$$\mathcal{F}_s = \sigma(Z(u) : u \leq s).$$

Additionally, let $\mathcal{H}_s \subseteq \mathcal{A}$ be another sigma-algebra that describes external information, e.g. portfolio information, demographic trends, and so on. While $\mathcal{F}_s \vee \mathcal{H}_s$ represents the maximally available information at current time s , the insurer potentially uses only a subset of this information, here described by a sub-sigma-algebra

$$\mathcal{G}_s \subseteq \mathcal{F}_s \vee \mathcal{H}_s.$$

There are various reasons for reducing the information $\mathcal{F}_s \vee \mathcal{H}_s$, including the following motives:

- reducing the numerical complexity of actuarial calculations,
- a lack of data for estimating $\mathbb{P}|_{\mathcal{F}_s \vee \mathcal{H}_s}$,
- anti-discrimination laws and data privacy regulations.

Let us focus for the moment on information reductions of the individual information \mathcal{F}_s only. Any choice of \mathcal{G}_s between \mathcal{H}_s and $\mathcal{H}_s \vee \mathcal{F}_s$ can be reasonable. In the calculation of technical reserves for the insurer's balance sheet, it largely suffices to study portfolio averages only, so it can be appropriate to cut \mathcal{G}_s down to \mathcal{H}_s . On the other hand, for the calculation of solvency reserves where the insurer aims to get a complete picture of the risk situation, it is rather preferable to use the maximally available information $\mathcal{H}_s \vee \mathcal{F}_s$. A frequently used intermediate case is $\mathcal{G}_s = \mathcal{H}_s \vee \sigma(Z(s))$ since this corresponds to a Markov assumption for Z . Yet, in many insurance applications Z is actually not Markov, and then $\mathcal{G}_s = \mathcal{H}_s \vee \sigma(Z(s))$ constitutes an *as-if-Markov* approach. The latter approach often results from a lack of data for the statistical estimation of the full distribution of Z . Anti-discrimination laws restrict for example the use of the sex of the insured as a risk factor in premiums and surrender values. By deleting the information on the sex of the insured from \mathcal{G}_s , we obtain gender-neutral values. On the other hand, anti-discrimination laws can imply significant lapse risk, so the sex of the insured should be accounted for in solvency calculations.

4 Differential notation

Suppose that X is a right-continuous and non-decreasing real process. Let H be a jointly measurable process. Then we define a *stochastic integral*

$$\int_I H(u) dX(u) \tag{3}$$

on intervals $I \subseteq \mathbb{R}$ by pathwise Lebesgue–Stieltjes integration. By using the additive relation $dX(t) = dX^+(t) - dX^-(t)$, we expand the domain of the stochastic integral to processes of the form $X = X^+ - X^-$, where X^+ and X^- are non-decreasing, non-negative, right-continuous real processes that are chosen minimally. Suppose that

$$Y(t) - Y(s) = \int_{(s,t]} H(u) dX(u), \quad \forall (s, t] \subseteq I, \tag{4}$$

almost surely for some interval I . Then we write this fact briefly as

$$dY(t) = H(t) dX(t), \quad t \in I, \tag{5}$$

and call (5) a *stochastic differential equation*. If the paths of X are differentiable with derivative x , then (5) is equivalent to the pathwise ordinary differential equation

$$\frac{d}{dt} Y(t) = H(t) x(t), \quad t \in I. \tag{6}$$

If X is a jump process with jumps at integer times n only, then (5) means that

$$Y(n) - Y(n - 1) = H(n) (X(n) - X(n - 1)), \quad (n - 1, n] \subseteq I. \tag{7}$$

5 The insurance cash flow

Suppose that B^b and B^p are stochastic processes that describe at each time $t \geq 0$ the accumulated benefits and accumulated premiums of an individual insurance policy on the interval $[0, t]$. The right-continuous process

$$B = B^b - B^p \tag{8}$$

is the total *insurance cash flow*. Note that $B(0-)$ is zero, but $B(0)$ can be different zero, describing a lump sum payment at time 0. Time t can either represent the contract time or the age of the insured, whichever is more convenient. If parameter t represents the contract time, then the insured has a positive starting age x at time zero, but we are not showing parameter x in our notation.

Definition 5.1 We say that B has a *deterministic canonical cash flow representation* if there exist non-decreasing, non-negative and right-continuous real functions $(B_i^+)_i$, $(B_i^-)_i$ and measurable and bounded real functions $(b_{ij})_{ij:i \neq j}$ that almost surely satisfy

$$dB(t) = \sum_i I_i(t-) dB_i(t) + \sum_{i,j:i \neq j} b_{ij}(t) dN_{ij}(t), \quad t \geq 0, \tag{9}$$

for $B_i := B_i^+ - B_i^-$. We say that a canonical cash flow representation has a *finite horizon* if there exists a time $T < \infty$ such that

$$dB_i^+(t) = 0, \quad dB_i^-(t) = 0, \quad b_{ij}(t) = 0, \quad t > T, \quad i, j \in \mathcal{Z}, \quad i \neq j.$$

We interpret $B_i^+(t)$ and $B_i^-(t)$ as accumulated sojourn benefits and accumulated premiums on $[0, t]$ in state i . The function $b_{ij}(t)$ describes the transition payment for a jump from i to j at time t .

There are only few insurance cash flows in insurance practice that can not be represented as deterministic canonical cash flows, provided that we allow Z to be non-Markov. Markov models struggle with duration dependencies in the insurance benefits such as deferment periods. For the modelling of a deferment period, it is actually not necessary to model the full duration process, but it suffices to introduce sub-states, say i_0 and i_1 , that indicate whether the deferment period has been completed yet. The insured jumps into sub-state i_0 first and then moves at completion of the deferment period to i_1 with probability 1. As such a splitting of states implies duration effects, we cannot continue to assume that the state process Z is Markov. Yet, if we generally drop the Markov assumption, then we are free to use the splitting method at discretion.

Let κ be a strictly positive and measurable real function that describes the value of the insurers investment portfolio at time t . Looking from the perspective of the current time s , the discounted accumulated future payments Y^+ and the discounted accumulated past and present payments Y^- of the insurance contract are given by

$$\begin{aligned} Y^+ &= \int_{(s, \infty)} \frac{\kappa(s)}{\kappa(t)} dB(t), \\ Y^- &= \int_{[0, s]} \frac{\kappa(s)}{\kappa(t)} dB(t). \end{aligned} \tag{10}$$

6 State occupation probabilities and transition rates

Definition 6.1 The almost surely unique right-continuous processes $(P_i(t))_{t \geq 0}$ that satisfy

$$P_i(t) = \mathbb{E}[I_i(t) | \mathcal{G}_s], \quad t \geq 0,$$

are called the *state occupation processes* with respect to information \mathcal{G}_s .

An explicit right-continuous definition of P_i can be obtained by calculating the above conditional expectations for each time t on the basis of a fixed regular conditional probability $\mathbb{P}(\cdot | \mathcal{G}_s)$, since then the right-continuity of P_i is a result of the right-continuity of I_i and the dominated convergence theorem. Right-continuous processes are generally almost surely unique.

Let $(P_{ij})_{ij}$ be the almost surely unique right-continuous processes that satisfy

$$P_{ij}(t) = \mathbb{E}[N_{ij}(t) | \mathcal{G}_s], \quad t \geq 0.$$

Definition 6.2 Let $(\Lambda_{ij})_{ij:i \neq j}$ be the right-continuous processes defined by $\Lambda_{ij}(s) = 0$ and

$$\begin{aligned} d\Lambda_{ij}(t) &= \frac{\mathbb{1}_{\{P_i(t-) > 0\}}}{P_i(t-)} dP_{ij}(t), \quad t > s, \\ d\Lambda_{ij}(t) &= \frac{\mathbb{1}_{\{P_j(t) > 0\}}}{P_j(t)} dP_{ij}(t), \quad t \leq s. \end{aligned} \tag{11}$$

We denote $d\Lambda_{ij}(t), t > s$, as *forward transition rate* and $d\Lambda_{ij}(t), t < s$, as *backward transition rate* for a jump from i to j .

The real-valued stochastic processes $(\Lambda_{ij})_{ij:i \neq j}$ are well-defined on $[0, T]$ if the integrability condition

$$\int_{(0,s]} \frac{\mathbb{1}_{\{P_j(t) > 0\}}}{P_j(t)} dP_{ij}(t) + \int_{(s,T]} \frac{\mathbb{1}_{\{P_i(t-) > 0\}}}{P_i(t-)} dP_{ij}(t) < \infty \tag{12}$$

holds on Ω for each $i, j \in \mathcal{Z}$ with $i \neq j$. In the remaining paper we generally assume that (12) is indeed satisfied. Note that one can drop Assumption (12) by redefining $(\Lambda_{ij})_{ij:i \neq j}$ as random measures on the real line, but for the sake of simplicity this paper prefers to see $(\Lambda_{ij})_{ij:i \neq j}$ as real-valued stochastic processes.

For each $i \in \mathcal{Z}$ we set

$$\begin{aligned} \Lambda_{ii}(t) &:= - \sum_{j:j \neq i} \Lambda_{ij}(t), \quad t > s, \\ \Lambda_{ii}(t) &:= - \sum_{j:j \neq i} \Lambda_{ji}(t), \quad t \leq s. \end{aligned} \tag{13}$$

Theorem 6.3 *The processes $(P_i)_i$ and $(\Lambda_{ij})_{ij:i \neq j}$ almost surely satisfy the stochastic differential equations*

$$dP_i(t) = \sum_j P_j(t-) d\Lambda_{ji}(t), \quad t > s, \tag{14}$$

$$dP_i(t) = - \sum_j P_j(t) d\Lambda_{ij}(t), \quad t \leq s. \tag{15}$$

For the proof see [Appendix](#).

Example 6.4 (Discrete-time Markov model) Suppose that Z is a Markov process that jumps only at integer times n . Let $\mathcal{G}_s = \sigma(Z(s))$ and $s \in \mathbb{N}_0$. Then $\Lambda_{ij}, i \neq j$, are pure jump processes with jumps at integer times of size

$$\begin{aligned} \Delta\Lambda_{ij}(n+1) &= \mathbb{P}(Z(n+1) = j | Z(n) = i) =: {}_1p_n^{ij}, \quad n \geq s, \\ \Delta\Lambda_{ij}(n-1) &= \mathbb{P}(Z(n-1) = i | Z(n) = j) =: {}_{-1}p_n^{ij}, \quad n \leq s. \end{aligned}$$

By using the fact that

$$P_i(n) = \sum_k I_k(s) {}_{n-s}p_s^{ki}, \quad n \geq 0, \tag{16}$$

for ${}_{n-s}p_s^{ki}$ defined as

$${}_{n-s}p_s^{ki} := \mathbb{P}(Z(n) = i | Z(s) = k), \quad n \geq 0,$$

Equation (14) is equivalent to the deterministic recursion equations

$${}_{n+1-s}p_s^{ki} = \sum_j {}_{n-s}p_s^{kj} {}_1p_n^{ji}, \quad n \geq s, k \in \mathcal{Z}.$$

The latter formula is known as the *Chapman–Kolmogorov equation*. Analogously, Eq. (15) corresponds to

$${}_{n-1-s}p_s^{ki} = \sum_j {}_{n-s}p_s^{kj} {}_{-1}p_n^{ji}, \quad n \leq s, k \in \mathcal{Z}.$$

Example 6.5 (Discrete-time as-if-Markov model) We start with the same setting as in Example 6.4 but drop the assumption that Z is Markov. Then Eq. (14) is equivalent to

$${}_{n+1-s}p_s^{ki} = \sum_j {}_{n-s}p_s^{kj} {}_1p_n^{kji}, \quad n \geq s,$$

for ${}_1p_n^{kji}$ defined as

$${}_1p_n^{kji} := \mathbb{P}(Z(n+1) = i | Z(n) = j, Z(s) = k). \tag{17}$$

Different from the Chapman–Kolmogorov equation in Example 6.4, the annual transition probability ${}_1p_n^{kji}$ really needs the extra parameter k here. Equation (14) is equivalent to

$${}_{n-1-s}p_s^{ki} = \sum_j {}_{n-s}p_s^{kj} {}_{-1}p_n^{kji}, \quad n \leq s,$$

for ${}_{-1}p_n^{kji}$ defined as

$${}_{-1}p_n^{kji} := \mathbb{P}(Z(n-1) = i | Z(n) = j, Z(s) = k). \tag{18}$$

Example 6.6 (Continuous-time Markov model) Suppose that Z is a Markov process. Let $\mathcal{G}_s = \sigma(Z(s))$, and assume that Λ_{ij} is continuously differentiable with derivative λ_{ij} . Then one can show that

$$\lambda_{ij}(t) = \frac{d}{dh} \Big|_{h=0} p_{ij}(t, t+h), \quad t > s,$$

for $i \neq j$, where the mappings

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

are known as *transition probabilities*. Because of

$$P_i(t) = \sum_k I_k(s) p_{ki}(s, t), \quad s, t \geq 0, \tag{19}$$

Equation (14) is equivalent to the ordinary differential equations

$$\frac{d}{dt} p_{ki}(s, t) = \sum_j p_{kj}(s, t) \lambda_{ji}(t), \quad t > s, k \in \mathcal{Z}. \tag{20}$$

The latter equation is known as the *Kolmogorov forward equation*. Analogously, Eq. (14) corresponds to the ordinary differential equations

$$\frac{d}{dt} p_{ki}(s, t) = - \sum_j p_{kj}(s, t) \lambda_{ij}(t), \quad t \leq s, k \in \mathcal{Z}. \tag{21}$$

The latter equation is not (!) the Kolmogorov backward equation.

Example 6.7 (Continuous-time as-if-Markov model) We start with the same setting as in Example 6.6 but drop the assumption that Z is Markov. Equation (19) still holds here, but Eq. (14) takes the form

$$\frac{d}{dt} p_{ki}(s, t) = \sum_j p_{kj}(s, t) \lambda_{k,ji}(t), \quad t > s, k \in \mathcal{Z}, \tag{22}$$

where $\lambda_{k,ij}(t) dt = d\Lambda_{k,ij}(t)$ is defined as in Definition 6.2 but with P_i and P_{ij} replaced by

$$P_{k,i}(t) = \mathbb{E}[I_i(t) | Z(s) = k], \quad P_{k,ij}(t) = \mathbb{E}[N_{ij}(t) | Z(s) = k].$$

Different from the Markov case, the transition rate $\lambda_{k,ji}$ needs the extra parameter k here. However, for the special case of decrement models Buchardt et al. [2] show the existence of artificial transition rates that are constant in k and still satisfy (20). Equation (14) takes here the form

$$\frac{d}{dt} p_{ki}(s, t) = - \sum_j p_{kj}(s, t) \lambda_{k,ij}(t), \quad t \leq s, k \in \mathcal{Z}. \tag{23}$$

Example 6.8 (Continuous-time doubly-stochastic Markov model) We start with the same setting as in Example 6.7 but expand the current information at time s to $\mathcal{G}_s = \sigma(Z(s), X(s))$, where X is a process that generates the external information \mathcal{H} . We assume that Z is conditionally Markov given X . This setup is known as

doubly-stochastic Markov model and is assumed in Norberg [9] and Buchardt [1]. Equation (14) has then the form

$$\frac{d}{dt}P_{(k,x),i}(s, t) = \sum_j P_{(k,x),j}(s, t) \lambda_{(k,x),ij}(t), \quad t > s, k \in \mathcal{Z}, \tag{24}$$

where $\lambda_{(k,x),ij}(t)$ is defined as in Definition 6.2 but with P_i and P_{ij} replaced by

$$P_{(k,x),i}(t) = \mathbb{E}[I_i(t)|(Z(s), X(s)) = (k, x)], \quad P_{k,ij}(t) = \mathbb{E}[N_{ij}(t)|(Z(s), X(s)) = (k, x)].$$

Equation (24) is equal to equation (4.1) in Buchardt [1], and our forward transition rates $\lambda_{(k,x),ij}(t)$ are mathematically equivalent to formula (39) in Norberg [9] and Definition 4.1 in Buchardt [1]. Our definition of forward transition rates is mathematically more rigorous and can help to clarify the definitions of Norberg [9] and Buchardt [1]. Theorem 6.3 shows that for Eq. (24) to hold it is actually not necessary to assume a doubly-stochastic Markov structure.

7 Statistical estimation of transition rates

This section demonstrates how to estimate the transition rates $(d\Lambda_{ij})_{i \neq j}$ from empirical data. The results presented here serve as a proof of concept and do not cover the full range of possible model settings. We recommend to additionally consult the statistical literature on landmark estimators.

A random variable ζ that generates the information \mathcal{G}_s is denoted as *landmark*. We aim to estimate

$$\begin{aligned} \Lambda_{z,ij}(t) &= \int_{(s,t]} \frac{\mathbb{1}_{P_{z,i}(u-) > 0}}{P_{z,i}(u-)} dP_{z,ij}(u), \quad t > s, \\ \Lambda_{z,ij}(t) &= \int_{(t,s]} \frac{\mathbb{1}_{P_{z,j}(u) > 0}}{P_{z,j}(u)} dP_{z,ij}(u), \quad t \leq s, \end{aligned}$$

where $P_{z,i}$ and $P_{z,ij}$ are defined by

$$P_{z,i}(t) = \mathbb{E}[I_i(t)|\zeta = z], \quad P_{z,ij}(t) = \mathbb{E}[N_{ij}(t)|\zeta = z].$$

Suppose that we observe a sample of $n \in \mathbb{N}$ individuals

$$(Z^1(t))_{L^1 < t \leq R^1}, \dots, (Z^n(t))_{L^n < t \leq R^n},$$

where $L^m, R^m, m \in \{1, \dots, n\}$, are random variables that describe left-truncation and right-censoring in the data. By $I^m, N_{ij}^m, \zeta^m, m \in \{1, \dots, n\}$, we denote the indicator processes, counting processes and landmarks of each observed individual. Let $J^m, m \in \{1, \dots, n\}$, be Bernoulli random variables that are only then non-zero if ζ^m is $\sigma(Z^m(t) : L^m < t \leq C^m)$ -measurable, i.e. in case of $J^m = 1$ the landmark ζ^m can be indeed observed in the data. Let

$$\hat{N}_{z,ij}(t) = \sum_{m=1}^n \mathbb{1}_{\{J^m=1\}} \mathbb{1}_{\{\zeta^m=z\}} (N_{ij}^m(t \wedge R^m) - N_{ij}^m(t \wedge L^m)), \quad t \geq 0,$$

be the counting processes of the sub-sample that selects individuals whose landmarks are observable and equal to z , and let

$$\hat{L}_{z,i}(t) = \sum_{m=1}^n \mathbb{1}_{\{J^m=1\}} \mathbb{1}_{\{\zeta^m=z\}} \mathbb{1}_{\{L^m < t \leq R^m\}} I_i^m(t), \quad t \geq 0,$$

be the corresponding at-risk processes.

Definition 7.1 The *landmark Nelson–Aalen estimator* for $(\Lambda_{z,ij})_{ij:i \neq j}$ is defined as

$$\begin{aligned} \hat{\Lambda}_{z,ij}(t) &= \int_{(s,t]} \frac{\mathbb{1}_{\hat{L}_{z,i}(u-) > 0}}{\hat{L}_{z,i}(u-)} d\hat{N}_{z,ij}(u), \quad t > s, \\ \hat{\Lambda}_{z,ij}(t) &= \int_{(t,s]} \frac{\mathbb{1}_{\hat{L}_{z,i}(u) > 0}}{\hat{L}_{z,i}(u)} d\hat{N}_{z,ij}(u), \quad t \leq s. \end{aligned}$$

The landmark Nelson–Aalen estimator for $t > s$ was first introduced by Putter and Spitoni [12], who consider the landmark $\zeta = Z(s)$ only.

Definition 7.2 The *landmark Aalen–Johansen estimator* for $(P_{z,i})_i$ is defined as the solution of the differential equation systems

$$\begin{aligned} d\hat{P}_{z,i}(t) &= \sum_j \hat{P}_{z,j}(t-) d\hat{\Lambda}_{z,ji}(t), \quad t > s, \\ d\hat{P}_{z,i}(t) &= - \sum_j \hat{P}_{z,j}(t) d\hat{\Lambda}_{z,ij}(t), \quad t \leq s, \end{aligned} \tag{25}$$

with initial/terminal values

$$\hat{P}_{z,i}(s) = \frac{\hat{L}_{z,i}(s)}{\sum_i \hat{L}_{z,i}(s)}.$$

The following theorem offers sufficient conditions for the consistency of the landmark Nelson–Aalen estimator. Our focus is on models with internal information only. We use the proof of Overgaard [11]. Compared to the existing literature, we allow for a larger class of landmarks, and we add the time-reversed perspective.

Theorem 7.3 *Let $T < \infty$. Suppose that*

- (a) $(Z^1, \zeta^1, L^1, R^1, J^1), \dots, (Z^n, \zeta^n, L^n, R^n, J^n)$ are independent and identically distributed,
- (b) R^1, L^1 are stochastically independent of $(Z^1(t))_{t \geq 0}, \zeta^1$,

- (c) ζ^1 is $\sigma(Z^1(u) : u \in I)$ -measurable for some interval $I \subseteq [0, s]$ and $J^1 = \mathbb{1}_{\{I \subseteq [L^1, R^1]\}}$,
- (d) $\mathbb{P}(J^1 = 1, L^1 < t \leq R^1) \geq \varepsilon > 0$ for all $t \in (0, T]$.

Then for each $i, j \in \mathcal{Z}, i \neq j$, and each z with $\mathbb{P}(\zeta = z) > 0$ we have

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\hat{\Lambda}_{z,ij}(t) - \Lambda_{z,ij}(t)| \right] \rightarrow 0, \quad n \rightarrow \infty.$$

The proof is given in [Appendix](#). The next theorem shows consistency also for the landmark Aalen–Johansen estimator.

Theorem 7.4 *Suppose that the assumptions of Theorem 7.3 are satisfied. Then for each $i, j \in \mathcal{Z}, i \neq j$, and each z with $\mathbb{P}(\zeta = z) > 0$ we have*

$$\mathbb{E} \left[\sup_{t \in [0, T]} |\hat{P}_{z,i}(t) - P_{z,i}(t)| \right] \rightarrow 0, \quad n \rightarrow \infty.$$

The proof is given in [Appendix](#).

Example 7.5 (Discrete-time Markov model) The classical Nelson–Aalen estimator for ${}_1p_n^{k,ji}$ is

$$\frac{\sum_m \mathbb{1}_{\{L^m < n, n+1 \leq R^m\}} I_j^m(n) I_i^m(n+1)}{\sum_m \mathbb{1}_{\{L^m < n, n+1 \leq R^m\}} I_j^m(n)}.$$

Plugging these estimators into the recursion equations in [Example 6.4](#) (Chapman–Kolmogorov equation) and solving the latter with the initial value [\(26\)](#) yields the classical Aalen–Johansen estimator.

Example 7.6 (Discrete-time as-if-Markov model) We start from the setting of [Example 7.5](#) but drop the assumption that Z is Markov. Let $\zeta := Z(s)$. Then the landmark Nelson–Aalen estimator for ${}_1p_n^{k,ji}$ in [\(17\)](#) is

$$\frac{\sum_m \mathbb{1}_{\{L^m < s, n+1 \leq R^m\}} I_k^m(s) I_j^m(n) I_i^m(n+1)}{\sum_m \mathbb{1}_{\{L^m < s, n+1 \leq R^m\}} I_k^m(s) I_j^m(n)},$$

and the landmark Nelson–Aalen estimator for ${}_{-1}p_n^{k,ji}$ in [\(18\)](#) is

$$\frac{\sum_m \mathbb{1}_{\{L^m < n-1, s \leq R^m\}} I_k^m(s) I_j^m(n) I_i^m(n-1)}{\sum_m \mathbb{1}_{\{L^m < n-1, s \leq R^m\}} I_k^m(s) I_j^m(n)}.$$

If we plug these estimators into the recursion equations in [Example \(6.5\)](#) and solve them with the initial/terminal values

$$\frac{\sum_m \mathbb{1}_{\{L^m < s \leq R\}} I_j^m(s)}{\sum_i \sum_m \mathbb{1}_{\{L^m < s \leq R\}} I_i^m(n)}, \quad j \in \mathcal{Z}, \tag{26}$$

then the solution is just the landmark Aalen–Johansen estimator $\hat{P}_{z,i}$, which converges here to the transition probability $\mathbb{P}(Z(t) = i | Z(s) = z)$. The landmarking ensures that we always estimate the transition probabilities consistently even if the data is non-Markov. In return, we increase the variance of the estimators since the landmarking uses sub-samples only.

Example 7.7 (Select mortality tables) Consider a discrete-time insurance model and let ζ represent the number of years that the insured already spent in the current state. Then the annual transition rates $({}_1p_s^{z,ij})_{z,ij}$ are known as *select tables at age s*. The corresponding landmark Nelson–Aalen estimator

$$\frac{\sum_m \mathbb{1}_{\{L^m < s, s+1 \leq R^m\}} \mathbb{1}_{\{\zeta=z\}} I^m j i(s) I_i^m(s+1)}{\sum_m \mathbb{1}_{\{L^m < s, s+1 \leq R^m\}} \mathbb{1}_{\{\zeta=z\}} I_j^m(s)}$$

equals the common raw estimate for ${}_1p_s^{z,ij}$.

Example 7.8 (Continuous-time as-if-Markov model) The landmark Nelson–Aalen estimator $\hat{\Lambda}_{z,ij}$ is a jump process, so a derivative $\hat{\lambda}_{z,ij}$ does not exist. This fact is completely analogous to the Markov case with the classical Nelson–Aalen estimator. Densities $\hat{\lambda}_{z,ij}$ can be obtained with the help of additional smoothing techniques. Without such smoothing steps our estimated model is in fact a discrete-time model (on an irregular time grid).

Remark 7.9 (Semi-Markov model) Let $U = (U(t))_{t \geq 0}$ be the duration process of Z , which is defined as

$$U(t) := \inf \{u \leq t : Z(t) = Z(r) \forall r \in [u, t]\}.$$

The state process Z is called a semi-Markov process if the bivariate process (Z, U) is Markov. In this case, $\zeta = (Z(s), U(s))$ is a natural choice for the landmark. Yet, $U(s)$ is usually not a discrete random variable, so we struggle with the fact that ζ takes values from an uncountably infinite set. A possible way out is to approximate ζ by a discrete random variable, e.g. replace $U(s)$ in ζ by the rounded duration $[hU(s)]/h$ for a suitably small $h > 0$.

8 Prospective and retrospective reserves

The discounted accumulated future payments Y^+ and the discounted accumulated past and present payments Y^- of the insurance contract, cf. (10), are in general not adapted to the current information \mathcal{G}_s . This motivates the following definition

Definition 8.1 The *prospective reserve* and *retrospective reserve* at time s are defined as

$$\begin{aligned} V^+ &= \mathbb{E}[Y^+ | \mathcal{G}_s], \\ V^- &= \mathbb{E}[Y^- | \mathcal{G}_s]. \end{aligned} \tag{27}$$

Either the prospective reserve or the retrospective reserve is commonly credited to the policyholder. The reserves typically serve as surrender values upon lapse of the policy, and they are used for identifying surplus and losses. Therefore, the choice of the sub-sigma-algebra \mathcal{G}_s affects the risk transfer between the individual insured, the insurance portfolio and the insurer. If we choose $\mathcal{G}_s = \mathcal{H}_s$, then V^+ and V^- describe mean portfolio reserves since all individual risk is averaged out. If we set $\mathcal{G}_s = \mathcal{H}_s \vee \mathcal{F}_s$, then V^+ and V^- describe fully individual reserves, in particular we have $V^- = Y^-$. If we define $\mathcal{G}_s = \mathcal{H}_s \vee \sigma(Z(s))$, then V^+ and V^- describe partially averaged reserves, but in case that Z is Markov the prospective reserve is still fully individual.

Theorem 8.2 *Suppose that the insurance cash flow B has a deterministic canonical representation of the form (9) with finite horizon $T < \infty$. Then*

$$V^+ = \sum_i \int_{(s,T]} \frac{\kappa(s)}{\kappa(t)} P_i(t-) dB_i(t) + \sum_{i,j:i \neq j} \int_{(s,T]} \frac{\kappa(s)}{\kappa(t)} b_{ij}(t) P_i(t-) d\Lambda_{ij}(t), \tag{28}$$

$$V^- = \sum_i \int_{[0,s]} \frac{\kappa(s)}{\kappa(t)} P_i(t-) dB_i(t) + \sum_{i,j:i \neq j} \int_{[0,s]} \frac{\kappa(s)}{\kappa(t)} b_{ij}(t) P_j(t) d\Lambda_{ij}(t), \tag{29}$$

almost surely.

The proof is given in [Appendix](#). Let

$$V_z^+ := \mathbb{E}[Y^+ | \zeta = z], \quad V_z^- := \mathbb{E}[Y^- | \zeta = z].$$

Theorem 8.3 *Suppose that the assumptions of Theorem 7.3 are satisfied. Let \hat{V}_z^+ and \hat{V}_z^- be the solutions of Eqs. (28) and (29) but with P_i and Λ_{ij} replaced by their landmark Nelson–Aalen and landmark Aalen–Johansen estimators. Then for each $i \in \mathcal{Z}$ and z with $\mathbb{P}(\zeta = z) > 0$ we have*

$$\mathbb{E}[|\hat{V}_z^+ - V_z^+|] \rightarrow 0, \quad \mathbb{E}[|\hat{V}_z^- - V_z^-|] \rightarrow 0, \quad n \rightarrow \infty.$$

The proof is given in [Appendix](#).

By combining Theorems 6.3 and (8.2), we can construct numerical schemes for the calculation of V^+ and V^- . By solving the stochastic differential equations

$$\begin{aligned}
 dW^+(t) &= \sum_i \frac{\kappa(s)}{\kappa(t)} P_i(t-) dB_i(t) + \sum_{i,j:i \neq j} \frac{\kappa(s)}{\kappa(t)} b_{ij}(t) P_i(t-) d\Lambda_{ij}(t), \\
 dP_i(t) &= \sum_j P_j(t-) d\Lambda_{ji}(t), \quad i \in \mathcal{Z},
 \end{aligned}
 \tag{30}$$

pathwise on $(s, T]$ with initial values of $W^+(s) = 0$ and $P_i(s) = \mathbb{E}[I_i(s)|\mathcal{G}_s]$, $i \in \mathcal{Z}$, we obtain the prospective reserve as

$$V^+ = W^+(T).$$

Similarly, by solving the stochastic differential equations

$$\begin{aligned}
 dW^-(t) &= - \sum_i \frac{\kappa(s)}{\kappa(t)} P_i(t-) dB_i(t) - \sum_{i,j:i \neq j} \frac{\kappa(s)}{\kappa(t)} b_{ij}(t) P_j(t) d\Lambda_{ij}(t), \\
 dP_i(t) &= - \sum_j P_j(t) d\Lambda_{ij}(t), \quad i \in \mathcal{Z},
 \end{aligned}
 \tag{31}$$

pathwise on $[0, s]$ with terminal values of $W^-(s) = 0$ and $P_i(s) = \mathbb{E}[I_i(s)|\mathcal{G}_s]$, $i \in \mathcal{Z}$, we obtain the retrospective reserve as

$$V^- = W^-(0) + B(0).$$

Example 8.4 (Discrete-time as-if-Markov model) Suppose that we are in the setting of Example 6.5. Let $T \in \mathbb{N}$ and $\kappa(s)/\kappa(t) = v^{t-s}$, i.e. we have a constant annual discounting factor of v . Let $(B_i)_i$ be step functions with jumps of size $b_i(n)$ at integer times n . Then (30) is equivalent to the recursion equations

$$\begin{aligned}
 W_k^+(n+1) &= W_k^+(n) + \sum_i v^{n+1-s} {}_{n-s}p_s^{ki} \left(b_i(n+1) + \sum_{j:j \neq i} b_{ij}(n+1) {}_1p_n^{kj} \right), \\
 {}_{n+1-s}p_s^{ki} &= \sum_j {}_{n-s}p_s^{kj} {}_1p_n^{ki}, \quad n \geq s, \\
 {}_0p_s^{ki} &= \delta_{ki}, \quad W^+(s) = 0.
 \end{aligned}$$

By calculating the recursion for each state k , starting from s going up to time T , we obtain the prospective reserve as

$$V^+ = W_{Z(s)}^+(T).$$

Likewise, (31) is equivalent to the backward recursion equations

$$\begin{aligned}
 W_k^-(n-1) &= W_k^-(n) + \sum_i v^{n-1-s} p_{n-s}^{ki} \left(b_i(n-1) + \sum_{j:j \neq i} b_{ji}(n-1) {}_{-1}p_n^{kji} \right), \\
 {}_{n-1-s}p_s^{ki} &= \sum_j p_{n-s}^{kj} {}_{-1}p_n^{kij}, \quad n \leq s, \\
 {}_0p_s^{ki} &= \delta_{ki}, \quad W^-(s) = 0.
 \end{aligned}$$

By calculating the recursion for each state k , starting from s and going down to time 0, we obtain the retrospective reserve as

$$V^- = W_{Z(s)}^-(0) + B(0).$$

Example 8.5 (Discrete-time Markov model) If we assume in Example 8.4 that Z is a Markov process, then the recursion formulas are the same but the parameter k in ${}_1p_n^{kij}$ and ${}_{-1}p_n^{kij}$ can be neglected.

Example 8.6 (Continuous-time as-if-Markov model) Suppose that we are in the setting of Example 6.7. Furthermore, let $(B_i)_i$ be continuously differentiable with derivatives $(b_i)_i$. Recall the definition of $p_{ij}(s, t)$ in Example 6.6. Solving (30) pathwise means here that we have to distinguish between the events $\{Z(s) = k\}$, $k \in \mathcal{Z}$. So we have to solve the ordinary differential equation system

$$\begin{aligned}
 \frac{d}{dt} W_k^+(t) &= \sum_i \frac{\kappa(s)}{\kappa(t)} p_{ki}(s, t) \left(b_i(t) + \sum_{j:j \neq i} b_{ij}(t) \lambda_{kij}(t) \right), \\
 \frac{d}{dt} p_{ki}(s, t) &= \sum_j p_{kj}(s, t) \lambda_{kji}(t), \\
 W_k^+(s) &= 0, \quad p_{ki}(s, s) = \delta_{ki}
 \end{aligned} \tag{32}$$

on $(s, T]$ for $k, i \in \mathcal{Z}$, and then we obtain the prospective reserve as

$$V^+ = W_{Z(s)}^+(T).$$

Solving (31) pathwise means that we need to solve the differential equation system

$$\begin{aligned}
 \frac{d}{dt} W_k^-(t) &= - \sum_i \frac{\kappa(s)}{\kappa(t)} p_{ki}(s, t) \left(b_i(t) + \sum_{j:j \neq i} b_{ji}(t) \lambda_{kji}(t) \right), \\
 \frac{d}{dt} p_{ki}(s, t) &= - \sum_j p_{kj}(s, t) \lambda_{kij}(t), \\
 W_k^-(s) &= 0, \quad p_{ki}(s, s) = \delta_{ki}
 \end{aligned} \tag{33}$$

on $[0, s)$ for each $k \in \mathcal{Z}$, and then we obtain the retrospective reserve as

$$V^- = W_{Z(s)}^-(0).$$

Example 8.7 (Continuous-time Markov model) If we assume in Example 8.6 that Z is a Markov process, then the differential equation systems are the same but the

parameter k in $\lambda_{k,ij}$ can be neglected. In the special case of decrement models with sojourn payments only and no transition payments, Buchardt et al. [2] showed that the parameter k may be even dropped if Z is non-Markov, but the transition rates have to be chosen in a specific artificial way. .

Example 8.8 (Continuous-time doubly-stochastic Markov model) Suppose that we are in the setting of Example 6.8. Let the differentiability assumptions of Example 8.6 be satisfied. Then Eq. (30) takes the form

$$\begin{aligned} \frac{d}{dt} W_{(k,x)}^+(t) &= \sum_i \frac{\kappa(s)}{\kappa(t)} P_{(k,x),i}(s, t) \left(b_i(t) + \sum_{j:j \neq i} b_{ij}(t) \lambda_{(k,x),ij}(t) \right), \\ \frac{d}{dt} P_{(k,x),i}(s, t) &= \sum_j P_{(k,x),j}(s, t) \lambda_{(k,x),ji}(t), \\ W_{(k,x)}^+(s) &= 0, \quad P_{(k,x),i}(s, s) = \delta_{ki}, \end{aligned} \tag{34}$$

and we obtain the prospective reserve as

$$V^+ = W_{(Z(s),X(s))}^+(T).$$

Equation (34) is equivalent to equations (5.2) & (4.1) in Buchardt [1]. Note that Eq. (34) does not really need the doubly-stochastic Markov structure.

Example 8.9 (Time-continuous semi-Markov model) Let Z be a semi-Markov process with corresponding duration process U , cf. Example 7.9. Let $\mathcal{G}_s = \sigma(Z(s), U(s))$. Suppose that $(B_i)_i$ and $(\Lambda_{ij})_{ij:i \neq j}$ are continuously differentiable with derivatives $(b_i)_i$ and $(\lambda_{ij})_{ij:i \neq j}$. Then we have

$$\begin{aligned} W^+(t) &= W_\zeta^+(t), \\ W^-(t) &= W_\zeta^-(t), \\ \lambda_{ij}(t) &= \lambda_{\zeta,ij}(t), \\ P_i(t) &= P_{\zeta,i}(t), \end{aligned}$$

for $\zeta := (Z(s), U(s))$ and suitable deterministic mappings $W_z^+, W_z^-, \lambda_{z,ij}, P_{z,i}$ because of the semi-Markov assumption. Solving (30) pathwise means here that we have to distinguish between the events $\{\zeta = z\}, z \in \mathcal{Z} \times [0, s]$. So we have to solve the ordinary differential equation system

$$\begin{aligned} \frac{d}{dt} W_z^+(t) &= \sum_i \frac{\kappa(s)}{\kappa(t)} P_{z,i}(t) \left(b_i(t) + \sum_{j:j \neq i} b_{ij}(t) \lambda_{z,ij}(t) \right), \\ \frac{d}{dt} P_i(t, z) &= \sum_j P_{z,j}(t) \lambda_{z,ji}(t), \\ W_z^+(s) &= 0, \quad P_{z,i}(s) = \delta_{iz_1} \end{aligned}$$

for each $z = (z_1, z_2) \in \mathcal{Z} \times [0, s] \times \mathbb{N}_0$, and then we obtain the prospective reserve as

$$V^+ = W_\zeta^+(T).$$

Solving (31) pathwise means that we need to solve the ordinary differential equation system

$$\begin{aligned} \frac{d}{dt} W_z^-(t) &= - \sum_i \frac{\kappa(s)}{\kappa(t)} P_{z,i}(t) \left(b_i(t) + \sum_{j:j \neq i} b_{ji}(t) \lambda_{z,ji}(t) \right), \\ \frac{d}{dt} P_{z,i}(t) &= - \sum_j P_{z,j}(t) \lambda_{z,ij}(t), \\ W_z^-(s) &= 0, \quad P_{z,i}(s) = \delta_{iz_1} \end{aligned}$$

for each $z = (z_1, z_2) \in \mathcal{Z} \times [0, s]$, and then we obtain the retrospective reserve as

$$V^- = W_\zeta^-(0).$$

As z is an element from $\mathcal{Z} \times [0, s]$, our ordinary differential equation systems involve an uncountably infinite number of equations. For numerical solving we need to reduce that system, for example by approximating $U(s)$ as described in Remark 7.9.

9 Conclusion

It is often unclear in insurance practice whether the true state process Z of an individual insured is actually a Markov process. We can then choose from the following options:

- (a) Pretend that Z is Markov and use the classical formulas for the calculation of reserves.
- (b) Replace Markov modelling by as-if-Markov evaluation as explained in this paper.

Option (a) comes with systematic model risk, which does not vanish for large sample sizes. The approximation error is difficult to quantify. The actuary might not even be able to tell whether the reserves have been overestimated or underestimated. Option (b) comes with additional unsystematic estimation risk, since the landmark estimators use sub-samples only, but the estimation risk vanishes for large sample sizes. Moreover, the statistical literature offers various methods for quantifying the estimation risk. This makes option (b) the preferable choice.

We demonstrated that the as-if-Markov transition rates that we need for option (b) can be well-estimated by the landmark Nelson–Aalen estimator. The statistical literature offers further results that wait to be utilized in insurance.

While this paper focusses on snapshots of prospective and retrospective reserves at fixed time points only, in actuarial risk management it is also

important to understand the dynamics of the reserves when time is moving. A comprehensive non-Markov theory for the dynamic perspective does not exist yet, but partial answers can be found in Christiansen and Furrer [3].

Appendix: Proofs

Proof of Theorem 6.3 Let $\mathbb{P}_\omega(\cdot)$ be a regular version of the conditional distribution $\mathbb{P}(\cdot | \mathcal{G}_s)(\omega)$. Let $\mathbb{E}_\omega[\cdot]$ be defined as the Lebesgue integral of the argument with respect to $\mathbb{P}_\omega(\cdot)$. Since $\mathbb{E}_\omega[I_i(t-)] = 0$ implies that $I_i(t-) = 0$ $\mathbb{P}_\omega(\cdot)$ -almost surely, by applying the Campbell theorem we obtain

$$\begin{aligned} \int_{(s,t]} \mathbb{1}_{\{P_i(u-)(\omega) > 0\}} dP_{ij}(u)(\omega) &= \mathbb{E}_\omega \left[\int_{(s,t]} \mathbb{1}_{\{P_i(u-)(\omega) > 0\}} I_i(u-) dN_{ij}(u) \right] \\ &= \mathbb{E}_\omega \left[\int_{(s,t]} I_i(u-) dN_{ij}(u) \right] \\ &= \int_{(s,t]} dP_{ij}(u)(\omega), \quad t > s, \end{aligned} \tag{35}$$

for \mathbb{P} -almost all $\omega \in \Omega$. Moreover, because of (2) we almost surely have

$$P_i(t) - P_i(s) = \sum_{j:j \neq i} (P_{ji}(t) - P_{ji}(s)) - \sum_{j:j \neq i} (P_{ij}(t) - P_{ij}(s)), \quad t > s.$$

By applying (35) and the definitions of Λ_{ij} and Λ_{ii} , the latter equation can be equivalently rewritten to

$$\begin{aligned} dP_i(t) &= \sum_{j:j \neq i} \mathbb{1}_{\{P_j(t-) > 0\}} dP_{ji}(t) - \sum_{j:j \neq i} \mathbb{1}_{\{P_i(t-) > 0\}} dP_{ij}(t) \\ &= \sum_j P_j(t-) d\Lambda_{ji}(t), \quad t > s. \end{aligned}$$

This shows (14). The proof of (15) is analogous. □

Proof of Theorem 7.3 Let $\mathbb{P}_z^c(\cdot) := \mathbb{P}(\cdot | \zeta = z)$ and $\mathbb{E}_z[\cdot] := \mathbb{E}[\cdot | \zeta = z]$. Let P_i^z and Λ_{ij}^z be the state occupation probabilities and the transition rates that correspond to the censored indicator process

$$I_i^z(t) = \mathbb{1}_{\{\zeta=z, J=1, L < t \leq R\}} I_i(t)$$

and the censored counting process

$$N_{ij}^z(t) = \mathbb{1}_{\{\zeta=z, J=1\}} (N_{ij}(t \wedge R) - N_{ij}(t \wedge L)) = \int_{(0,t]} \mathbb{1}_{\{\zeta=z, J=1, L < u \leq R\}} dN_{ij}(u).$$

By using assumption (b) and the Campbell Theorem, for $t \leq s$ we can show that

$$\begin{aligned} P_{ij}^z(t) - P_{ij}^z(s) &= \mathbb{E}[N_{ij}^z(t) - N_{ij}^z(s)] \\ &= \mathbb{E}\left[-\mathbb{1}_{\{\zeta=z\}} \int_{(t,s]} \mathbb{E}[\mathbb{1}_{\{J=1,L<u \leq R\}} | (Z_t)_{t \geq 0}, \zeta] dN_{ij}(u)\right] \\ &= -\mathbb{P}(\zeta = z) \mathbb{E}_z \left[\int_{(t,s]} \mathbb{E}[\mathbb{1}_{\{J=1,L<u \leq R\}}] dN_{ij}(u) \right] \\ &= -\mathbb{P}(\zeta = z) \int_{(t,s]} \mathbb{E}[\mathbb{1}_{\{J=1,L<u \leq R\}}] dP_{z,ij}(u). \end{aligned}$$

On the other hand, assumption (b) implies that

$$P_i^z(t-) = \mathbb{E}[I_i^z(t-)] = \mathbb{P}(\zeta = z) \mathbb{E}_z[I_i(t-)] \mathbb{E}[\mathbb{1}_{\{J=1,L<t \leq R\}}].$$

The latter two equations and assumption (d) yield

$$\Lambda_{z,ij}(t) = \Lambda_{ij}^z(t) \tag{36}$$

for $t \leq s$. Similar calculations show that (36) holds also for $t > s$. For $p \in [1, 2)$, we define the p -variation norm as $\|\cdot\|_{[p]} := \|\cdot\|_\infty + \|\cdot\|_{V_p}$, where $\|\cdot\|_\infty$ is the supremum norm on $[0, T]$ and $\|\cdot\|_{V_p}$ is the total p -variation on $[0, T]$. According to Theorem 3 in Overgaard [10], it holds that

$$\mathbb{E}\left[\|n^{-1}\hat{N}_{z,ij} - P_{ij}^z\|_{[p]}\right] \rightarrow 0, \quad n \rightarrow \infty, \tag{37}$$

for $p \in (1, 2)$, where $P_{ij}^z(t) := \mathbb{E}[N_{ij}^z(t)]$, $t \geq 0$. Because of Eq. (2), we have

$$\begin{aligned} &|n^{-1}\hat{L}_{z,i}(t) - P_i^z(t)| \\ &\leq |n^{-1}\hat{L}_{z,i}(s) - P_i^z(s)| + \sum_{i,j:i \neq j} |n^{-1}\hat{N}_{z,ij}(t) - P_{ij}^z(t)| + \sum_{i,j:i \neq j} |n^{-1}\hat{N}_{z,ij}(s) - P_{ij}^z(s)|. \end{aligned}$$

By using assumption (a), the law of large numbers, dominated convergence and Theorem 3 from Overgaard [10], we obtain

$$\mathbb{E}\left[\|n^{-1}\hat{L}_{z,i} - P_i^z\|_{[p]}\right] \rightarrow 0, \quad n \rightarrow \infty, \tag{38}$$

for $p \in (1, 2)$. The inequalities $\|\int_{(0,\cdot]} g(s)df(s)\|_{[p]} \leq k_p \|f\|_{[p]} \|g\|_{[p]}$, see Dudley [5], and $\|fg\|_{[p]} \leq \|f\|_{[p]} \|g\|_{[p]}$ imply that

$$\left\| \int_{(0,\cdot]} (g(s))^{-1} df(s) - \int_{(0,\cdot]} (g'(s))^{-1} df'(s) \right\|_{[p]} \leq \frac{\|g - g'\|_{[p]}}{\|gg'\|_{[p]}} \|f\|_{[p]} + \|g'\|_{[p]} \|f - f'\|_{[p]}. \tag{39}$$

Because of this inequality and (37), (38) and assumption (d), we can conclude that

$$\mathbb{E}\left[\|\hat{\Lambda}_{z,ij} - \Lambda_{ij}^z\|_{[p]}\right] \rightarrow 0, \quad n \rightarrow \infty, \tag{40}$$

for $p \in (1, 2)$. Finally, in the latter formula we replace Λ_{ij}^z by $\Lambda_{z,ij}$, see (36). □

Proof of Theorem 7.4 Let $\hat{P}_z(t) = (P_{z,i}(t))_i$ be a row vector and $\hat{\Lambda}_z(t) = (\hat{\Lambda}_{z,ij}(t))_{ij}$ a matrix, where the diagonal entries $\hat{\Lambda}_{z,ii}$ of the matrix are defined as in (13). The solution $\hat{P}_z(t)$ of (25) equals the product integrals

$$\begin{aligned} \hat{P}_z(t) &= \hat{P}_z(s) \prod_{(s,t]} (\mathbb{1} + d\hat{\Lambda}_{z,ji}), \quad t > s, \\ \hat{P}_z(t)^\top &= \prod_{(t,s]} (\mathbb{1} + d\hat{\Lambda}_{z,ji}) \hat{P}_z(t)^\top, \quad t \leq s, \end{aligned}$$

see Gill and Johansen [6]. Product integration is a continuous functional with respect to the supremum norm for sequences with uniformly bounded total 1-variation, see Gill and Johansen [6]. Therefore the assertion follows from Theorem 7.3 and Theorem 3 in Overgaard [10], which says that the expectation in (37) and, hence, the expectation in (40) are uniformly bounded for $p = 1$. □

Proof of Theorem 8.2 Note that the random variables Y^+ and Y^- have finite expectation since the boundedness of $(B_i^+)_i, (B_i^-)_i$ and $(b_{ij})_{ij:i \neq j}$ on $[0, T]$ implies

$$\max\{Y^+, Y^-\} \leq \sum_i (B_i^+(T) + B_i^-(T)) \sup_{0 \leq s \leq T} \frac{\kappa(s)}{\kappa(s)} \leq C \left(2 + \sum_{i,j:i \neq j} N_{ij}(T) \right)$$

for a constant $C < \infty$, and since (1) implies that $\sum_{i,j:i \neq j} N_{ij}(T)$ has finite expectation. By applying the conditional expectation $\mathbb{E}[\cdot | \mathcal{G}_s]$ on the sojourn payment parts of Y^+ and Y^- and pulling the conditional expectations inside the integrals, we directly obtain the sojourn payment parts of V^+ and V^- in (28) and (29). By applying the conditional expectation $\mathbb{E}[\cdot | \mathcal{G}_s]$ on the transition payment parts of Y^+ and Y^- and using the Campbell Theorem, we end up with the transition payment parts of V^+ and V^- according to formulas (28) and (29). □

Proof of Theorem 8.3 The formulas (28) and (29) are of the form $\int_{(0,1]} g(s)df(s)$. Analogously to (39) we can show that

$$\left\| \int_{(0,1]} g(s)df(s) - \int_{(0,1]} g'(s)df'(s) \right\|_{[p]} \leq \|g - g'\|_{[p]} \|f\|_{[p]} + \|g'\|_{[p]} \|f - f'\|_{[p]}.$$

Because of this inequality and (37), (38) and assumption (d), we can conclude that

$$\mathbb{E} \left[\|\hat{V}_z - V_z\|_{[p]} \right] \rightarrow 0, \quad n \rightarrow \infty.$$

□

Acknowledgements I would like to thank Christian Furrer for fruitful discussions and insightful remarks that helped to improve the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Buchardt K (2017) Kolmogorov's forward PIDE and forward transition rates in life insurance. *Scand Actuar J* 5:377–394
2. Buchardt K, Furrer C, Steffensen M (2019) Forward transition rates. *Financ Stochast* 23(4):975–999
3. Christiansen MC, Furrer C (2021) Dynamics of state-wise prospective reserves in the presence of non-monotone information. *Insur Math Econ* 97:81–98
4. Datta S, Satten GA (2001) Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models. *Stat Probab Lett* 55(4):403–411
5. Dudley RM (1992) Frechet differentiability, p-variation and uniform Donsker classes. *Ann Probab* 20(4):1968–1982
6. Gill R, Johansen S (1990) A survey of product-integration with a view towards application in survival analysis. *Ann Stat* 18(4):1501–1555
7. Guibert Q, Planchet F (2018) Non-parametric inference of transition probabilities based on Aalen–Johansen integral estimators for acyclic multi-state models: application to LTC insurance. *Insur Math Econ* 82:21–36
8. Niessl A, Allignol A, Müller C, Beyersmann J (2020) Estimating state occupation and transition probabilities in non-Markov multi-state models subject to both random left-truncation and right-censoring. [arXiv:2004.06514v1](https://arxiv.org/abs/2004.06514v1)
9. Norberg R (2010) Forward mortality and other vital rates—are they the way forward? *Insur Math Econ* 47(2):105–112
10. Overgaard M (2019a) Counting processes in p-variation with applications to recurrent events. [arXiv:1903.04296v1](https://arxiv.org/abs/1903.04296v1)
11. Overgaard M (2019b) State occupation probabilities in non-Markov models. *Math Methods Stat* 28(4):279–290
12. Putter H, Spitoni C (2018) Non-parametric estimation of transition probabilities in non-Markov multi-state models: the landmark Aalen–Johansen estimator. *Stat Methods Med Res* 27(7):2081–2092

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.