STEP SEMI-MARKOV MODELS AND APPLICATION TO MANPOWER MANAGEMENT

Vlad Stefan Barbu[1], Guglielmo D'Amico[2], Raimondo Manca[3]
and Filippo Petroni[4]

Abstract. The purpose of this paper is to introduce a class of stochastic processes that we call step semi-Markov processes and to illustrate the modelling capacity of such processes in practical applications. The name of this process comes from the fact that we have a semi-Markov process and the transition between two states is done through several steps. We first introduce these models and the main quantities that characterize them. Then, we derive the recursive evolution equations for two-step semi-Markov processes. The interest of using this type of model is illustrated by means of an application in manpower planning.

1. Introduction

Semi-Markov processes are an important class of stochastic processes that have been widely studied and applied in several fields, see *e.g.* [1,16,19,21]. In this paper we introduce a class of stochastic processes, called step semi-Markov processes (SSMP) that generalizes ordinary semi-Markov processes. To be more specific, SSMP are semi-Markov processes for which an additional insight is brought: the sojourn time in a state before making a transition represents the addition of two or several times that correspond to different physical causes. A typical example in survival analysis would be a time that is the sum between the incubation time of a disease and the waiting time before a change of state occurs. Clearly, this type of phenomenon could be modelled through a classical multi-state model (semi-Markov or other), by introducing a new state for each different time (see [10]), but the interest of our approach comes from several points:

(i) It is a different modelling technique that can capture some features that are not taken into account by existing frameworks. As previously mentioned, the holding time in a state before moving to another state is formed by two times, of different natures. Another point is that the transition from one state to another

[1] Université de Rouen, Laboratoire de Mathématiques Raphaël Salem, UMR 6085, Avenue de l'Université, BP.12, 76801 Saint-Étienne-du-Rouvray, France. barbu@univ-rouen.fr

[2] Dipartimento di Farmacia, Università "G. d'Annunzio" di Chieti-Pescara, via dei Vestini 31, 66100 Chieti, Italy. g.damico@unich.it

[3] Dipartimento MEMOTEF, Università di Roma "La Sapienza", via del Castro Laurenziano, 9, 00161 Rome, Italy. raimondo.manca@uniroma1.it

[4] Dipartimento di Scienze Economiche ed Aziendali, Universitá di Cagliari, via Sant'Ignazio, 17, 09123 Cagliari, Italy. fpetroni@unica.it

depends on the first time spent in the state; note also that the second time depends on the length of the first time.

For instance, in Section 5 we consider an application in manpower management, where the states of the system represent the different positions of an employee in a company and the holding time in a state is the sum of two times: the time spent from the last promotion and the time the employee does a training for a potential upgrade.

(ii) The type of dependence considered in our model is more general than the one obtained by extending the state space of the process.

(iii) In some applications involving complex systems or networks, there is a natural interest to reduce the number of states to a minimum that is necessary, something that is achieved by the type of model that we propose.

In this paper we define SSMP and we discuss their relation with ordinary semi-Markov processes, the latter being a special case of the former. We study the evolution equations of the transition probability function of the model. We are interested in the general model where we take into account the initial backward process, as well as in the simplest model, without considering the initial backward. We show that, according to the existing relation between the value of the backward recurrence time process and the length of the first step of the transition, there are two different evolution equations. This is a characteristic of the two-step semi-Markov model that is not present for ordinary semi-Markov processes. Furthermore we describe how to get estimation of the parameters of a SSMP considering a sample path of a two-step semi-Markov model censored at a fixed arbitrary time $M \in \mathbb{N}^*$. The practical relevance of the advanced model is illustrated by means of an application in manpower management (MPM), framework that has been an important field of application of Markov and semi-Markov processes, [6, 7, 14, 17, 22–24, 27–30]. In the application we show the appropriateness of SSMP for MPM problems and especially for the efficient integration of the training of workers within a manpower planning model.

Our paper is structured as follows: in the next section we first introduce some classical semi-Markov notations and we define the step semi-Markov model. Section 3 is devoted to the study of the recursive evolution equations for step semi-Markov processes. After that, some elements of statistical estimation are presented. In Section 5 we explain how to apply a two-step model in MPM; in this context, an application and related results will be presented. Some remarks on future work will be given in the last section.

2. THE MODEL

Let us consider a random system with finite state space $E = \{1, \ldots, s\}$. Assume that the time evolution of the system is governed by a stochastic process $Z = (Z_k)_{k \in \mathbb{N}}$. Let us denote by $T = (T_n)_{n \in \mathbb{N}}$ the successive time points when state changes in $(Z_k)_{k \in \mathbb{N}}$ occur and by $J = (J_n)_{n \in \mathbb{N}}$ the successive visited states at these time points. Set also $X = (X_n)_{n \in \mathbb{N}}$ for the successive sojourn times in the visited states. Thus $X_n = T_n - T_{n-1}$, $n \in \mathbb{N}^*$, and, by convention, we set $X_0 = T_0 = 0$. The relation between the process Z and the process J of the successively visited states is given by $Z_k = J_{N(k)}$, or, equivalently, $J_n = Z_{T_n}, n, k \in \mathbb{N}$, where $N(k) := \max\{n \in \mathbb{N} \mid T_n \leq k\}$ is the discrete-time counting process of the number of jumps in $[1, k] \subset \mathbb{N}$.

We denote by $\mathcal{M}_E$ the set of real matrices on $E \times E$ and by $\mathcal{M}_E(\mathbb{N})$ the set of matrix-valued functions defined on $\mathbb{N}$, with values in $\mathcal{M}_E$. For $\mathbf{A} \in \mathcal{M}_E(\mathbb{N})$, we write $\mathbf{A} = (\mathbf{A}(k); \ k \in \mathbb{N})$, where, for $k \in \mathbb{N}$ fixed, $\mathbf{A}(k) = (A_{ij}(k); \ i, j \in E) \in \mathcal{M}_E$.

We assume that the chain $(J, T) = (J_n, T_n)_{n \in \mathbb{N}}$ is a *Markov renewal chain (MRC)* and that the chain $Z = (Z_k)_{k \in \mathbb{N}}$ is a *semi-Markov chain (SMC)* associated to the MRC (J, T), that is, for all $n \in \mathbb{N}$, for all $i, j \in E$, and for all $k \in \mathbb{N}$ it satisfies almost surely

$$\mathbb{P}(J_{n+1} = j, T_{n+1} - T_n = k \mid J_0, \ldots, J_n; T_0, \ldots, T_n) = \mathbb{P}(J_{n+1} = j, T_{n+1} - T_n = k \mid J_n). \qquad (2.1)$$

All through this paper we consider that equation (2.1) is independent of n, i.e. that (J, T) is *homogeneous*. Then, the discrete-time semi-Markov kernel $\tilde{\mathbf{q}}$ is defined by $\tilde{q}_{ij}(k) := \mathbb{P}(J_{n+1} = j, X_{n+1} = k \mid J_n = i)$. We denote

by $\tilde{\mathbf{p}} = (\tilde{p}_{ij})_{i,j \in E} \in \mathcal{M}_E$ the transition matrix of the Markov chain (J_n), defined by

$$\tilde{p}_{ij} := \mathbb{P}(J_{n+1} = j \mid J_n = i), \ i, j \in E, \ n \in \mathbb{N}.$$

Let us also introduce the sojourn time distributions in a given state and the conditional distributions depending on the next state to be visited.

Definition 2.1. For all $i, j \in E$, let us define:

(1) $\tilde{f}_{ij}(\cdot)$, the conditional distribution of X_{n+1}, $n \in \mathbb{N}$:

$$\tilde{f}_{ij}(k) = \mathbb{P}(X_{n+1} = k \mid J_n = i, J_{n+1} = j), \ k \in \mathbb{N}. \tag{2.2}$$

(2) $\tilde{F}_{ij}(\cdot)$, the conditional distribution function of X_{n+1}, $n \in \mathbb{N}$:

$$\tilde{F}_{ij}(k) = \mathbb{P}(X_{n+1} \le k \mid J_n = i, J_{n+1} = j) = \sum_{l=0}^{k} \tilde{f}_{ij}(l), \ k \in \mathbb{N}. \tag{2.3}$$

(3) $\tilde{h}_i(\cdot)$, the sojourn time distribution in state i:

$$\tilde{h}_i(k) = \mathbb{P}(X_{n+1} = k \mid J_n = i) = \sum_{j \in E} \tilde{q}_{ij}(k), \ k \in \mathbb{N}^*. \tag{2.4}$$

(4) $\tilde{H}_i(\cdot)$, the sojourn time distribution function in state i:

$$\tilde{H}_i(k) = \mathbb{P}(X_{n+1} \le k \mid J_n = i) = \sum_{l=1}^{k} \tilde{h}_i(l), \ k \in \mathbb{N}^*. \tag{2.5}$$

Note that, for all $i, j \in E$ and $k \in \mathbb{N}$ such that $\tilde{p}_{ij} \neq 0$, the semi-Markov kernel verifies the relation

$$\tilde{q}_{ij}(k) = \tilde{p}_{ij} \tilde{f}_{ij}(k).$$

For G the cumulative distribution function of a r.v. X, we denote its *survival function* by $\overline{G}(n) := 1 - G(n) = \mathbb{P}(X > n)$, $n \in \mathbb{N}$.

Let us now consider that the sojourn time X_{n+1} in state T_n before going to state T_{n+1} is the sum of two different times, say U_{n+1} and V_{n+1}, $X_{n+1} = U_{n+1} + V_{n+1}$. Consequently, the semi-Markov conditional hypothesis given in equation (2.1) can be written

$$\mathbb{P}(J_{n+1} = j, U_{n+1} = u, V_{n+1} = v \mid J_0, \dots, J_n; U_0, \dots, U_n, V_0, \dots, V_n)$$
$$= \mathbb{P}(J_{n+1} = j, U_{n+1} = u, V_{n+1} = v \mid J_n). \tag{2.6}$$

Consequently, the semi-Markov kernel can be written for any time $k \in \mathbb{N}$

$$\tilde{q}_{ij}(k) = \mathbb{P}(J_{n+1} = j, U_{n+1} + V_{n+1} = k \mid J_n = i)$$
$$= \sum_{u=0}^{k} \mathbb{P}(J_{n+1} = j, U_{n+1} + V_{n+1} = k, U_{n+1} = u \mid J_n = i)$$
$$= \sum_{u=0}^{k} \mathbb{P}(V_{n+1} = k - u \mid U_{n+1} = u, J_n = i, J_{n+1} = j)$$
$$\times \mathbb{P}(J_{n+1} = j \mid J_n = i, U_{n+1} = u) \mathbb{P}(U_{n+1} = u \mid J_n = i)$$
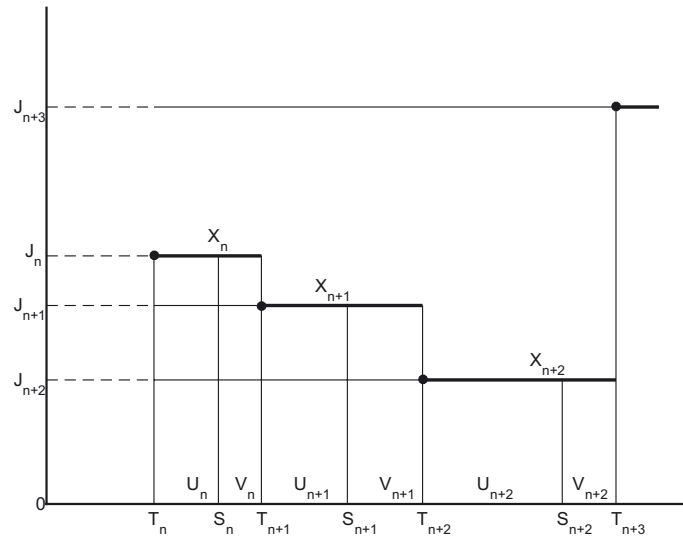$$= \sum_{u=0}^{k} p_{iu;j} f_{iu;j}(k - u) g_i(u), \tag{2.7}$$

FIGURE 1. A two-step semi-Markov chain trajectory.

where we used the following notation:

- $p_{iu;j} = \mathbb{P}(J_{n+1} = j \mid J_n = i, U_{n+1} = u)$,
- $f_{iu;j}(v) = \mathbb{P}(V_{n+1} = v \mid J_n = i, J_{n+1} = j, U_{n+1} = u)$,
- $g_i(u) = \mathbb{P}(U_{n+1} = u \mid J_n = i)$.

For all states $i, j \in E$ and positive integer $u \in \mathbb{N}$, note that $(g_i(k), k \in \mathbb{N})$ and $(f_{iu;j}(v), v \in \mathbb{N})$ are probability distributions, that is

$$\sum_{k \in \mathbb{N}} g_i(k) = 1 \text{ and } \sum_{v \in \mathbb{N}} f_{iu;j}(v) = 1.$$

Note also that for any positive integer $u \in \mathbb{N}$, $(p_{iu;j})_{i,j \in E}$ is a stochastic matrix, that is

$$\sum_{j \in \mathbb{E}} p_{iu;j} = 1.$$

Definition 2.2. A Markov renewal chain (J, T) whose semi-Markov kernel verifies equation (2.7) is called a *two-step Markov renewal chain* and the associated semi-Markov chain $Z = (Z_k)_{k \in \mathbb{N}}$ is called a *two-step semi-Markov chain*.

A trajectory of a two-step semi-Markov chain is illustrated in Figure 1.
   Several remarks are worth to be done here.

Remark 2.3. Note that a two-step Markov renewal chain (respectively a two-step semi-Markov chain) is just a usual Markov renewal chain (respectively a usual semi-Markov chain), endowed with a supplementary structure: the waiting times are the sum of two times (denoted here by U_n and V_n), representing two different types of randomness.

Remark 2.4. Although all the work presented in this paper is carried out in discrete time, note that the same type of results can be presented in continuous time.

Remark 2.5. Note also that there is a very simple and intuitive explanation behind the equation (2.7), describing the dynamics of the system from state i to state j, after a global holding time equal to $u + v$: first, the first waiting time in state i, $U_{n+1} = u$, is randomly chosen according to the probability distribution $(g_i(k), k \in \mathbb{N})$; then, the next state to be visited is randomly chosen, depending on this first time u, and this choice is done according to the Markov transition matrix $(p_{iu;j})_{i,j \in E}$; finally, the second waiting time in state i before going to state j is randomly chosen, depending on the first time u, and this choice is done according to the probability distribution $(f_{iu;j}(v), v \in \mathbb{N})$.

Let us also introduce the following notation:

- $q_{iu;j}(v) = \mathbb{P}(J_{n+1} = j, V_{n+1} = v \mid J_n = i, U_{n+1} = u) = p_{iu;j}f_{iu;j}(v)$,
- $h_{iu}(v) = \mathbb{P}(V_{n+1} = v \mid J_n = i, U_{n+1} = u) = \sum_{j \in E} f_{iu;j}(v)p_{iu;j}$,
- $H_{iu}(v) = \mathbb{P}(V_{n+1} \le v \mid J_n = i, U_{n+1} = u)$,
- $F_{iu;j}(v) = \mathbb{P}(V_{n+1} \le v \mid J_n = i, J_{n+1} = j, U_{n+1} = u)$,
- $G_i(u) = \mathbb{P}(U_{n+1} \le u \mid J_n = i)$.

As previously mentioned, for G the cumulative distribution function of a r.v., the associated survival function is denoted by $\overline{G}(n)$. Consequently, for all states $i, j \in E$ and positive integer $u \in \mathbb{N}$, we put $\overline{F}_{iu;j}$, $\overline{H}_{iu}$ and $\overline{G}_i$ for the corresponding survival functions.

Remark 2.6. Note that Definition 2.2 can immediately be generalized in order to define a $m-$step Markov renewal chain or $m-$step semi-Markov chain. On the one hand, the only thing to be considered is that the sojourn time X_{n+1} in state T_n before going to state T_{n+1} is the sum of m different times, say $U_{n+1}^1, \ldots, U_{n+1}^{m-1}$ and V_{n+1}, $X_{n+1} = U_{n+1}^1 + \ldots + U_{n+1}^{m-1} + V_{n+1}$. On the other hand, the interest (and the difficulty) of studying such models comes from the fact that the dependence between these different holding times has to be specified. For instance, we can write

$$\mathbb{P}(J_{n+1} = j, U_{n+1}^1 = u^1, \ldots, U_{n+1}^{m-1} = u^{m-1}, V_{n+1} = v \mid J_n = i)$$
$$= \mathbb{P}(J_{n+1} = j, V_{n+1} = v \mid J_n = i, U_{n+1}^1 = u^1, \ldots, U_{n+1}^{m-1} = u^{m-1})$$
$$\times \mathbb{P}(U_{n+1}^1 = u^1, \ldots, U_{n+1}^{m-1} = u^{m-1} \mid J_n = i).$$

Note that both quantities in the right hand term could be written in various ways, taking into account different types of dependence between the times $U_{n+1}^1, \ldots, U_{n+1}^{m-1}$ and V_{n+1}.

It is obvious that such a process would be more flexible than a two-step process and maybe more adapted for some applications. A possible drawback of such a model is that, as it is more complex, it is more difficult to work with, to estimate parameters, *etc.* Note also that one could consider processes with a different number of steps for each transition; possible situations could occur when this number m of steps depends on the present state or on the future state or on both; with the notation used in the previous equation, we could have $m = m_i$ or $m = m_j$ or $m = m_{ij}$, $i, j \in E$). This will be the object of a future work.

The relations between the characteristics of a semi-Markov process (semi-Markov kernel $\tilde{q}_{ij}(k)$, conditional holding time distributions $\tilde{f}_{ij}(k)$, Markov transition probability $\tilde{p}_{ij}$) and the corresponding quantities of a two-step semi-Markov process are given in the next proposition. The proof of these results is immediate.

Proposition 2.7. *For all states $i, j \in E$ and nonnegative $k \in \mathbb{N}$ we have:*

(1) $\tilde{p}_{ij} = \sum_{u \in \mathbb{N}} p_{iu;j}g_i(u)$,

(2) $\tilde{q}_{ij}(k) = \sum_{u=0}^{k} q_{iu;j}(k - u)g_i(u) = \sum_{u=0}^{k} p_{iu;j}f_{iu;j}(k - u)g_i(u)$,

(3) $\tilde{f}_{ij}(k) = \frac{\tilde{q}_{ij}(k)}{\tilde{p}_{ij}} = \frac{\sum_{u_2=0}^{k} p_{iu_2;j}f_{iu_2;j}(k-u_2)g_i(u_2)}{\sum_{u_1 \in \mathbb{N}} p_{iu_1;j}g_i(u_1)}$, *for $\tilde{p}_{ij} \ne 0$.*

The following remark, although quite obvious, will be needed in the sequel.

Remark 2.8. Note that $\mathbb{P}(U_{n+1} = u \mid J_n = i)$ in general is not equal to $\mathbb{P}(U_{n+1} = u \mid J_n = i, J_{n+1} = j), i, j \in E, u \in \mathbb{N}$. Intuitively, this fact is clear because there is a dependence between J_{n+1} and U_{n+1}. A direct computation of both terms yields the same conclusion.

3. EVOLUTION EQUATIONS

In this section, we will study the evolution equations of the transition probabilities of a two-step semi-Markov chain.

Let us first consider the case when we take into account the initial backward time of the semi-Markov process. The evolution of the semi-Markov process can be described by the semi-Markov transition probabilities that are defined as follows.

Definition 3.1. The transition function with initial backward process of the two-step semi-Markov chain Z_t is the matrix-valued function ${}^b\Phi = ({}^b\phi_{iu_1;j}(l;t); i,j \in E, u_1, l, t \in \mathbb{N}) \in \mathcal{M}_E$ defined by

$$
{}^b\phi_{iu_1;j}(l,t) := \mathbb{P}(Z_t = j \mid Z_0 = i, B_0 = l, U_1 = u_1), \tag{3.8}
$$

where the left upper-script b in ${}^b\phi$ stands for the initial backward and $B_t := t - T_{N(t)}$ is the backward time process associated to the semi-Markov process.

The following proposition establishes a recursive formula for computing the transition probability function ${}^b\Phi$ of the two-step semi-Markov chain.

Proposition 3.2. *For all states $i, j \in E$ and times $u_1, l, t \in \mathbb{N}$, such that $u_1 < l$ we have*

$$
\begin{aligned}
{}^b\phi_{iu_1;j}(l,t) = {} & \delta_{ij} \frac{\overline{H}_{iu_1}(t+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} \\
& + \sum_{m=1}^{t} \frac{q_{iu_1;j}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} \cdot \overline{G}_j(t-m-1) \\
& + \sum_{k \in E} \sum_{m=1}^{t} \sum_{u_2=1}^{t-m-1} \frac{q_{iu_1;k}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} \cdot g_k(u_2) \cdot {}^b\phi_{ku_2;j}(0,t-m).
\end{aligned} \tag{3.9}
$$

Similarly, if $u_1 \geq l$, then

$$
\begin{aligned}
{}^b\phi_{iu_1;j}(l,t) = {} & \delta_{ij} \overline{H}_{iu_1}(t+l-u_1) \\
& + \sum_{m=u_1-l}^{t} q_{iu_1;j}(m+l-u_1) \cdot \overline{G}_j(t-m-1) \\
& + \sum_{k \in E} \sum_{m=u_1-l}^{t} \sum_{u_2=1}^{t-m-1} q_{iu_1;k}(m+l-u_1) \cdot g_k(u_2) \cdot {}^b\phi_{ku_2;j}(0,t-m).
\end{aligned} \tag{3.10}
$$

The usual conventions $\overline{G}_j(-1) = 1$ and $\sum_{s=0}^{-1} = 0$ apply.

Proof. Let us consider first the case $u_1 < l$.

In general, the probability function can be represented as follows:

$$
\begin{aligned}
{}^b\phi_{iu_1;j}(l,t) := {} & \mathbb{P}(Z_t = j \mid Z_0 = i, B_0 = l, U_1 = u_1) \\
= {} & \mathbb{P}(Z_t = j, T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1) \tag{3.11} \\
& + \mathbb{P}(Z_t = j, T_1 \leq t \mid Z_0 = i, B_0 = l, U_1 = u_1). \tag{3.12}
\end{aligned}
$$

Now consider the first addendum (3.11) on the right hand side of the former equation for which the following chain of equalities holds true.

$$
\begin{aligned}
&\mathbb{P}(Z_t = j, T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1)\\
&= \mathbb{P}(Z_t = j \mid T_1 > t, Z_0 = i, B_0 = l, U_1 = u_1)\mathbb{P}(T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1)\\
&= \delta_{ij}\mathbb{P}(T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1)\\
&= \delta_{ij}\mathbb{P}(T_1 - T_0 > t - T_0 \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)\\
&= \delta_{ij}\mathbb{P}(U_1 + V_1 > t + l \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)\\
&= \delta_{ij}\mathbb{P}(V_1 > t + l - u_1 \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)\\
&= \delta_{ij}\frac{\mathbb{P}(V_1 > t + l - u_1, T_1 > 0 \mid J_0 = i, T_0 = -l, U_1 = u_1)}{\mathbb{P}(T_1 > 0 \mid J_0 = i, T_0 = -l, U_1 = u_1)}.
\end{aligned}
\tag{3.13}
$$

The numerator of (3.13) can be evaluated as follows:

$$
\begin{aligned}
\mathbb{P}(V_1 > t + l - u_1, T_1 > 0 \mid J_0 = i, T_0 = -l, U_1 = u_1) &= \mathbb{P}(V_1 > t + l - u_1, V_1 > l - u_1 \mid J_0 = i, T_0 = -l, U_1 = u_1)\\
&= \mathbb{P}(V_1 > t + l - u_1, \mid J_0 = i, U_1 = u_1) = \overline{H}_{iu_1}(t + l - u_1).
\end{aligned}
\tag{3.14}
$$

The denominator of (3.13) can be evaluated as follows:

$$
\begin{aligned}
\mathbb{P}(T_1 > 0 \mid J_0 = i, T_0 = -l, U_1 = u_1) &= \mathbb{P}(T_1 - T_0 > 0 - T_0 \mid J_0 = i, U_1 = u_1)\\
&= \mathbb{P}(U_1 + V_1 > l \mid J_0 = i, U_1 = u_1)\\
&= \mathbb{P}(V_1 > l - u_1 \mid J_0 = i, U_1 = u_1) = \overline{H}_{iu_1}(l - u_1).
\end{aligned}
\tag{3.15}
$$

This proves that

$$
\mathbb{P}(Z_t = j, T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1) = \delta_{ij}\frac{\overline{H}_{iu_1}(t + l - u_1)}{\overline{H}_{iu_1}(l - u_1)}.
\tag{3.16}
$$

Now, let us consider the second addendum (3.12).

$$
\begin{aligned}
&\mathbb{P}(Z_t = j, T_1 \le t \mid Z_0 = i, B_0 = l, U_1 = u_1)\\
&= \sum_{k \in E}\sum_{m=1}^{t}\sum_{u_2 \ge 0}\mathbb{P}(Z_t = j, T_1 = m, J_1 = k, U_2 = u_2 \mid Z_0 = i, B_0 = l, U_1 = u_1)\\
&= \sum_{k \in E}\sum_{m=1}^{t}\sum_{u_2 \ge 0}\mathbb{P}(Z_t = j \mid T_1 = m, J_1 = k, U_2 = u_2, J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)\\
&\quad \times \mathbb{P}(U_2 = u_2 \mid T_1 = m, J_1 = k)\mathbb{P}(T_1 = m, J_1 = k \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)\\
&= \sum_{k \in E}\sum_{m=1}^{t}\sum_{u_2 \ge 0}{}^{b}\phi_{ku_2;j}(0; t - m)\cdot g_k(u_2)\cdot\frac{q_{iu_1;k}(m + l - u_1)}{\overline{\overline{H}}_{iu_1}(l - u_1)}.
\end{aligned}
\tag{3.17}
$$

Formula (3.17) can be rewritten as follows:

$$
\sum_{k \in E}\sum_{m=1}^{t}\left(\sum_{u_2=0}^{t-m-1}{}^{b}\phi_{ku_2;j}(0; t - m)\cdot g_k(u_2)\cdot\frac{q_{iu_1;k}(m + l - u_1)}{\overline{\overline{H}}_{iu_1}(l - u_1)}\right.
$$

$$
\left. + \sum_{u_2 \ge t-m}{}^{b}\phi_{ku_2;j}(0; t - m)\cdot g_k(u_2)\cdot\frac{q_{iu_1;k}(m + l - u_1)}{\overline{H}_{iu_1}(l - u_1)}\right).
\tag{3.18}
$$

When $u_2 \geq t - m$, the probability ${}^b\phi_{ku_2;j}(0; t - m) = \delta_{kj}$ and, therefore, we have that

$$\sum_{u_2 \geq t-m} {}^b\phi_{ku_2;j}(0; t-m) \cdot g_k(u_2) \cdot \frac{q_{iu_1;k}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} = \sum_{u_2 \geq t-m} \delta_{kj} \cdot g_k(u_2) \cdot \frac{q_{iu_1;k}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)}, \tag{3.19}$$

which, in turn, implies that

$$\sum_{k \in E} \sum_{m=1}^{t} \sum_{u_2 \geq t-m} {}^b\phi_{ku_2;j}(0; t-m) \cdot g_k(u_2) \cdot \frac{q_{iu_1;k}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} = \sum_{m=1}^{t} \sum_{u_2 \geq t-m} \frac{q_{iu_1;j}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} \cdot g_j(u_2)$$

$$= \sum_{m=1}^{t} \frac{q_{iu_1;j}(m+l-u_1)}{\overline{H}_{iu_1}(l-u_1)} \cdot \overline{G}_j(t-m-1). \tag{3.20}$$

A substitution of (3.20) into (3.18) and the consideration of formula (3.16) leads to formula (3.9). Given the similarity with formula (3.9), we shall only sketch the proof of formula (3.10). We consider the usual representation of the probability function

$$
\begin{aligned}
{}^b\phi_{iu_1;j}(l,t) &= \mathbb{P}(Z_t = j, T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1) && \text{(3.21)} \\
&\quad + \mathbb{P}(Z_t = j, T_1 \leq t \mid Z_0 = i, B_0 = l, U_1 = u_1). && \text{(3.22)}
\end{aligned}
$$

Now, for $u_1 \geq l$, the computations are modified as follows.
First,

$$
\begin{aligned}
\mathbb{P}(Z_t = j, T_1 > t \mid Z_0 = i, B_0 = l, U_1 = u_1) &= \delta_{ij} \mathbb{P}(T_1 - T_0 > t - T_0 \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1) \\
&= \delta_{ij} \frac{\mathbb{P}(V_1 > t + l - u_1 \mid J_0 = i, U_1 = u_1)}{\mathbb{P}(T_1 > 0 \mid J_0 = i, T_0 = -l, U_1 = u_1)} \\
&= \delta_{ij} \frac{\overline{H}_{iu_1}(t + l - u_1)}{1} = \delta_{ij} \overline{H}_{iu_1}(t + l - u_1). && \text{(3.23)}
\end{aligned}
$$

$$\tag{3.24}$$

For $T_1 \leq t$ we have

$$\mathbb{P}(Z_t = j, T_1 \leq t \mid Z_0 = i, B_0 = l, U_1 = u_1)$$

$$= \sum_{k \in E} \sum_{m \geq u_1 - l}^{t} \sum_{u_2 \geq 0} \mathbb{P}(Z_t = j \mid T_1 = m, J_1 = k, U_2 = u_2, J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)$$

$$\times \mathbb{P}(U_2 = u_2 \mid T_1 = m, J_1 = k) \mathbb{P}(T_1 = m, J_1 = k \mid J_0 = i, T_0 = -l, T_1 > 0, U_1 = u_1)$$

$$= \sum_{k \in E} \sum_{m \geq u_1 - l}^{t} \sum_{u_2 \geq 0} {}^b\phi_{ku_2;j}(0; t-m) \cdot g_k(u_2) \cdot q_{iu_1;k}(m+l-u_1). \tag{3.25}$$

A decomposition of the sum $\sum_{u_2 \geq 0} = \sum_{u_2=0}^{t-m-1} + \sum_{u_2 \geq t-m}$ and the fact that for $u_2 \geq t - m$ the probability ${}^b\phi_{ku_2;j}(0; t-m) = \delta_{kj}$ yield equation (3.10), after some simple substitutions. $\qquad\square$

The transition function ${}^b\phi_{iu_1;j}(l,t)$ gives the probabilistic evolution of the system and is dependent on the backward process value, $B_0 = l$. An interesting case is obtained for $B_0 = 0$, that is, when the current time zero is a transition time. In this case, the transition probability function simplifies considerably. Indeed, note first of all that u_1 cannot be less than $l(= 0)$; note also that the first case ($u_1 < l$) makes no sense anymore. Second,

for $u_1 \geq l = 0$, we have

$$
\begin{aligned}
{}^b\phi_{iu_1;j}(0,t) := \phi_{iu_1;j}(t) = {} & \delta_{ij}\overline{H}_{iu_1}(t - u_1) \\
& + \sum_{m=u_1}^{t} q_{iu_1;j}(m - u_1) \cdot \overline{G}_j(t - m - 1) \\
& + \sum_{k \in E} \sum_{m=u_1}^{t} \sum_{u_2=1}^{t-m-1} q_{iu_1;k}(m - u_1) \cdot g_k(u_2) \cdot \phi_{ku_2;j}(0, t - m).
\end{aligned}
\tag{3.26}
$$

The evolution equations for the two-step semi-Markov chain model can be further generalized when taking into account the initial or final backward and forward processes as done for non-homogeneous semi-Markov chains (see [5]).

4. SOME ASSOCIATED ESTIMATION TOPICS

Starting from available data, one can obtain empirical estimators of the quantities of interest defining our model. Let us consider a sample path of a two-step semi-Markov model censored at a fixed arbitrary time $M \in \mathbb{N}^*$,

$$
J_0, U_1, V_1, J_1, U_2, V_2, \ldots, J_{N(M)-1}, U_{N(M)}, V_{N(M)}, J_{N(M)}, B_M,
\tag{4.27}
$$

where $N(M) := \max\{n \mid T_n \leq M\}$ is the counting process of the number of jumps in $[1; M]$ and $B_M := M - T_{N(M)}$ is the censored sojourn time in the last visited state $J_{N(M)}$.

For any states $i, j \in E$ and $u, v \in \mathbb{N}^*$, let us consider:

(1) $N_i(M) := \sum_{n=0}^{N(M)-1} \mathbf{1}_{\{J_n=i\}} = \sum_{n=0}^{M} \mathbf{1}_{\{J_n=i, T_{n+1}\leq M\}}$ - the number of visits to state i of the MC $(J_n)_{n\in\mathbb{N}}$, up to time M;

(2) $N_i(u, M) := \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i, U_n=u\}} = \sum_{n=1}^{M} \mathbf{1}_{\{J_{n-1}=i, U_n=u, T_n\leq M\}}$ - the number of visits to state i of the MC $(J_n)_{n\in\mathbb{N}}$, up to time M, with sojourn time $U = u$ in state i;

(3) $N_{ij}(u, M) := \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i, J_n=j, U_n=u\}} = \sum_{n=1}^{M} \mathbf{1}_{\{J_{n-1}=i, J_n=j, U_n=u, T_n\leq M\}}$ - the number of transitions of the MC $(J_n)_{n\in\mathbb{N}}$ from state i to state j, up to time M, with sojourn time $U = u$ in state i;

(4) $N_{ij}(u, v, M) := \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i, J_n=j, U_n=u, V_n=v\}}$
$= \sum_{n=1}^{M} \mathbf{1}_{\{J_{n-1}=i, J_n=j, U_n=u, V_n=v, T_n\leq M\}}$ - the number of transitions of the MC $(J_n)_{n\in\mathbb{N}}$ from state i to state j, up to time M, with sojourn time $U = u$ and $V = v$ in state i.

Consequently, the empirical estimators of the quantities of interest defining the model are given by:

$$
\widehat{p}_{iu;j}(M) := \frac{N_{ij}(u, M)}{N_i(u, M)};
\tag{4.28}
$$

$$
\widehat{g}_i(u, M) := \frac{N_i(u, M)}{N_i(M)};
\tag{4.29}
$$

$$
\widehat{f}_{iu;j}(v, M) := \frac{N_{ij}(u, v, M)}{N_{ij}(u, M)}.
\tag{4.30}
$$

From classical arguments, these estimators are strongly consistent, as $M \to \infty$.

If one considers several (say K) sample paths of a two-step semi-Markov model censored at a fixed arbitrary time $M \in \mathbb{N}^*$,

$$
J_0^l, U_1^l, V_1^l, J_1^l, U_2^l, V_2^l, \ldots, J_{N(M)-1}^l, U_{N(M)}^l, V_{N(M)}^l, J_{N(M)}^l, B_M^l, l = 1, \ldots, K,
$$

then the corresponding empirical estimators are given by

$$\widehat{p}_{iu;j}(K) := \frac{N_{ij}(u, K)}{N_i(u, K)},$$

$$\widehat{g}_i(u, K) := \frac{N_i(u, K)}{N_i(K)},$$

$$\widehat{f}_{iu;j}(v, K) := \frac{N_{ij}(u, v, K)}{N_{ij}(u, K)},$$

where the quantities $N_{ij}(u, K), N_i(u, K), N_{ij}(u, v, K), N_i(K)$ are defined taking into account the counting over all the K sample paths. As before, these estimators are strongly consistent, as $K \to \infty$.

At this point it is possible to estimate the probability $q_{iu;j}(v)$ through the estimator $\widehat{q}_{iu;j}(v, K) := \widehat{p}_{iu;j}(K)\widehat{f}_{iu;j}(v, K)$. The transition functions can be estimated by plugging-in $\widehat{p}_{iu;j}(K)$, $\widehat{f}_{iu;j}(v, K)$ and $\widehat{q}_{iu;j}(v, K)$ inside equations (3.9) and (3.10).

It is worth noticing that one can consider the estimation framework when the U_n's and V_n's are not directly observed, but the observations are $J_0, X_1, J_1, X_2, \ldots, J_{N(M)-1}, X_{N(M)}, J_{N(M)}, B_M$, where $X_n := U_n + V_n$ represent the successive holding times. This case with latent variables U_n and V_n can correspond to some important applied situations. In this case, estimating the (conditional) distribution of U_n and V_n could be a challenging topic, involving convolution estimation for dependent random variables. When considering general $m-$step semi-Markov chains instead of two-step semi-Markov ones, like in Remark 2.6, this problem of estimation could be even more complicated. This research direction represents the topic of a future work.

5. Application to manpower management

5.1. Manpower dynamics using two-step semi-Markov models

The application of our model tackles the problem of time evolution of salary lines and more precisely of the ongoings in the case of promotion. This problem is part of the larger problem of manpower planning. Manpower management (MPM) is an interdisciplinary topic. The literature is copious and the interested reader can refer to [4, 11] for older references and to [6, 8, 13] for newer references on this topic.

At the beginning of the seventies, Markov processes were applied to this topic (see for example [4, 20]). This approach was further developed in the following years; we can mention the contributions of [2, 12, 25]. A considerable improvement was realized by introducing semi-Markov processes, see *e.g.* [3, 18, 23, 24, 26]. Note also that in [18] and in [15] a generalization of semi-Markov processes was also presented to introduce the age and the seniority into the model.

Recently the important problem of integrating the modelling of the training of workers within a manpower planning model of a company has been considered in [9, 12] and extended in [10] by considering the so called augmented semi-Markov models.

In this subsection we show how natural is the integration of the training period in a manpower system modelled by a two-step semi-Markov model.

Let us consider a MPM that incorporates the need of the employees to attend a period of training before being upgraded. Let us assume that a worker has just entered state i of the organization (company). First, we can imagine that he/she will maintain this position until a random time selected by the distribution $g_i(\cdot)$. Second, after a random time u (realization of $g_i(\cdot)$), he/she starts a period of training for potential upgrade (change) of his/her position. Furthermore it may be reasonable to assume that:

– Next position is selected according to the probability distribution $p_{iu;\cdot}$ and therefore depends on the current position i and on how long the worker occupied the position i, that is, on the time u;
– After selecting the next position, say state j, he/she will transit into j after the completion of the necessary training that will have a random length sampled from $f_{iu;j}(\cdot)$. This may be the case because the necessary training can depend on the past experience of the worker (i, u), as well as on the new position j to be held.

TABLE 1. The total population subdivided by age and state.

Ages	I Empl	II Empl	I Func	II Func	Exec	tot age
≤24	27	0	0	0	0	27
25–29	429	1	1	0	0	431
30–34	1510	95	14	0	0	1619
35–39	1969	307	71	4	3	2354
40–44	1794	361	206	6	5	2372
45–49	2354	675	479	30	9	3727
50–54	2766	934	1085	106	41	4932
55–59	927	261	499	95	39	1821
≥60	138	27	61	20	19	265
tot deg	12 094	2661	2416	261	116	17 548
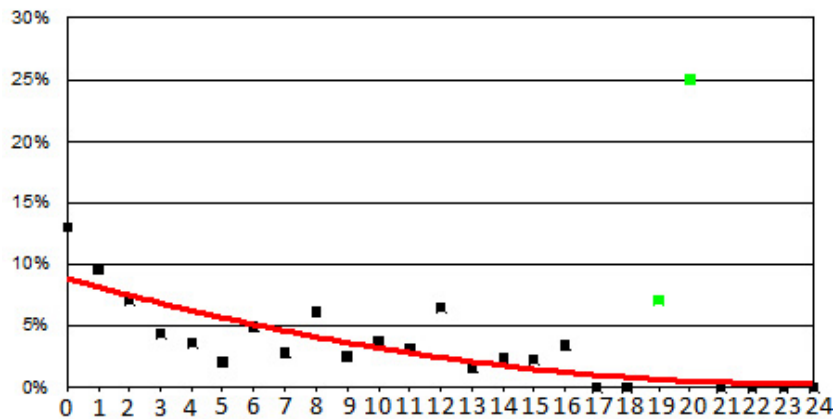
FIGURE 2. The probability to be promoted between two ranks of employees; the abscissa axis gives the seniority years.

Notice that, during the training period (second part of the sojourn time), the worker stays in state i because the upgrade is not yet completed and the company will pay a wage according to class i.

This description shows that the two-step semi-Markov chain model is a good candidate to study the career evolution of workers in a company and naturally can include among its main features the training period.

5.2. Data description

We have used an old dataset of a primary Italian bank. There were 15 ranks in the bank: 7 ranks of office workers, 5 of functionary workers and 3 of executives. The dataset contains only transitions among classes obtained by an aggregation of ranks into five states. The first state was formed by the first 5 ranks of office workers, the second one by the other two employee ranks, the third state was formed by the first four ranks of functionary workers. The last two states were formed by the last rank of functionary workers and by the three states of executives, respectively.

The total number of considered workers was exactly 17 548, out of which 5325 were females and 12 223 males. The initial distribution of the workers in the five states and their age structure is illustrated in Table 1.

We can also note that the probability to be promoted is higher with a lower seniority in the rank (time elapsed in the rank since the entrance in the rank) and decreases as the seniority is increasing. Figure 2 illustrates this phenomenon.
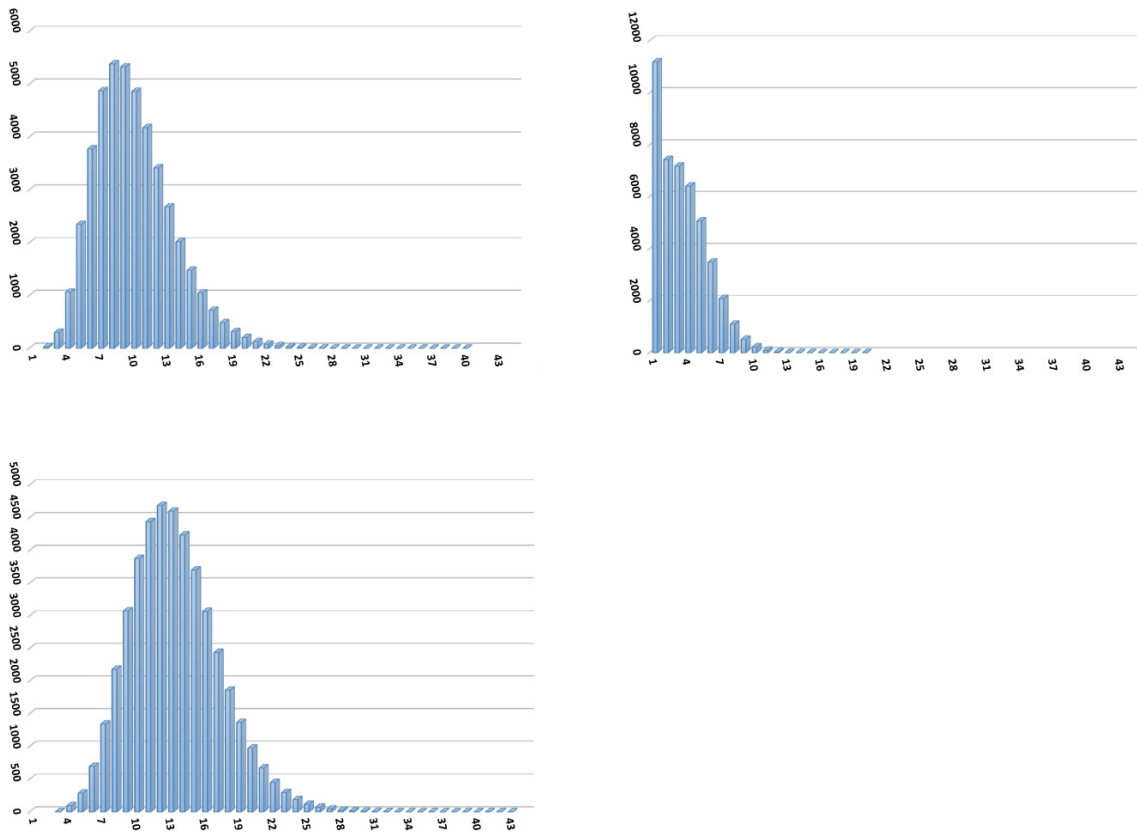
FIGURE 3. Histograms of values of U (*top-left*), V (*top-right*) and X (*bottom-left*).

TABLE 2. Descriptive statistics of pdf of U_n, V_n and X_n.

Variable	Mean	Standard deviation	Mode	Median	Pearson asymmetry
U	9.822	3.507	8	8	1.558
V	3.371	2.128	1	2	1.932
X	13.193	3.898	12	12	0.918

Our dataset is not complete for the application of the two-step semi-Markov model because neither did we have information on how the waiting time is divided in the two steps nor on whether the bank implemented a training scheme for the promotions. To overcome this problem we decided to work with a Monte Carlo simulation for the reconstruction of the subdivision of the waiting times into two steps. We worked in the following way. For each individual h in the dataset we dispose of the sequence of the waiting times $\{X_{n+1}^{(h)}\}$ for $n = 0, 1, ..., N^{(h)}$. We decided that the length of $U_{n+1}^{(h)}$ should be about three times the length of the time $V_{n+1}^{(h)}$ or equivalently, $U_{n+1}^{(h)}$ is, in average, the 75% of $X_{n+1}^{(h)}$ and $V_{n+1}^{(h)}$ the remaining 25%. Then given the total length of $X_{n+1}^{(h)}$ we did a pseudo-random number extraction from a Uniform distribution with support $[0, \frac{X_{n+1}^{(h)}}{2}] \cap \mathbb{N}$ and then we fixed $U_{n+1}^{(h)} = \frac{X_{n+1}^{(h)}}{2} + x$ where x is the result of the extraction. This operation is done for each waiting time of each individual.

In this way we reconstructed a dataset that contains all the necessary information for the implementation of the two-step semi-Markov model. In Figure 3 we report in the top-left corner the histogram of the absolute
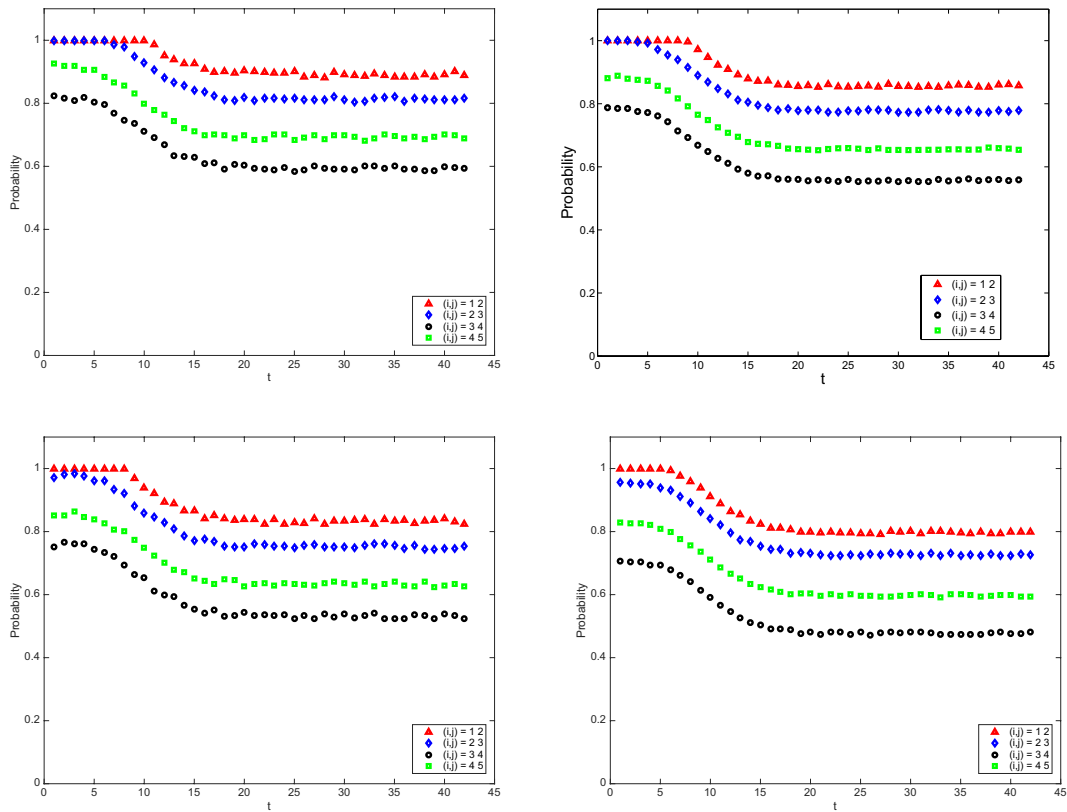
FIGURE 4. Probability function ${}^{b}\phi_{iu_1;j}(l,t)$ for $l = 0$ and $u_1 = 1$, as a function of time t, for different couples of states i and j. Average shares $\frac{U}{X}$ and $\frac{V}{X}$: $70-30\%$ (*top-left*), $75-25\%$ (*top-right*), $80-20\%$ (*bottom-left*), $85-15\%$ (*bottom-right*).

frequencies of the U variable; on the top-right corner the histogram is relative to the variable V and in the bottom-left corner the histogram of the waiting time X. In Table 1 we report some descriptive statistics of those histograms. As it is possible to see, the mean length of the waiting times X is 13 years and the mean value of the first step U is 9.822, so that U is 74.45% of X. The column of the standard deviations tells us that the variance of X is less than the sum of the variances of U and V. This is the case because in our simulation the variables U and V are negatively correlated because their sum is fixed by the value of X, hence, large values of U are associated with low values of V and *vice versa*. The three distributions show a positive asymmetry which is more pronounced for the second step V.

In order to carry out a sensitivity analysis, the simulation has been repeated not only dividing the waiting times X into first and second steps of shares $75-25\%$ but also in three other different proportions: $70-30\%$, $80-20\%$ and $85-15\%$.

By using the estimators presented in Section 4 we were able to have estimates of the main parameters of the model.

5.3. Results and discussion

In Figure 4 we show the behaviour of the probability ${}^{b}\phi_{iu_1;j}(l,t)$ for $l = 0$, $u_1 = 1$ as a function of time t for different couples of states i and j.
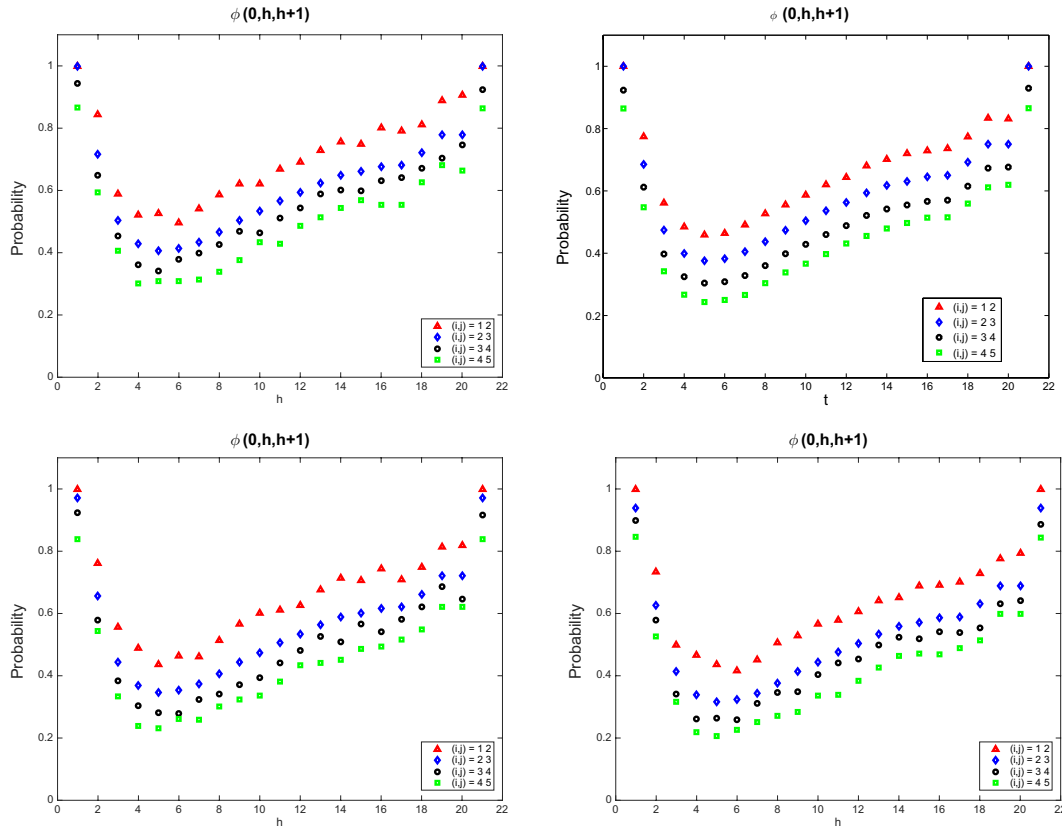
FIGURE 5. Probability function $^b\phi_{iu_1;j}(l,t)$ for $l = 0$, $u_1 = h$, $t = h + 1$, as a function of h, for different couples of states i and j. Average shares $\frac{U}{X}$ and $\frac{V}{X}$: $70-30\%$ (*top-left*), $75-25\%$ (*top-right*), $80-20\%$ (*bottom-left*), $85-15\%$ (*bottom-right*).

As we can see, in all the four cases these probabilities exhibit a decreasing behaviour for all couples of states. This means that if the worker entered state i at current time and after one unit of time he/she begins the training for possible promotion, it is more probable that he/she will be promoted within a short time. The passage from state 1 to state 2 shows more chance, in contrast to the promotion from state 3 to state 4, which is harder. It is also relevant to note that, as the percentage of time the individual is in the first step of the waiting time increases, the transition probability decreases. The differences are moderate because the total waiting time $X = U + V$ is kept fixed and also because the sharing of the waiting time in the two steps is done independently from the initial and final states of the waiting time.

In Figure 5 we show the behaviour of the probability $^b\phi_{iu_1;j}(l,t)$ for $l = 0$, $u_1 = h$, $t = h+1$ as a function of h for different couples of states i and j. For any considered couple of states i and j we see that this probability exhibits first a decreasing path and then it shows a reversion, with an increasing behaviour of the function. This means that if the worker entered state i at current time and after h units of time he/she begins the training, the probability to be promoted within one year is very high if the training starts anon, then decreases in time and then it starts to increase on the long time horizon. This tells us that, in our organization, the best workers (those starting a training in a short time) have more chances to be promoted compared to average workers. It is also evident that bad workers (those entering the training later than the others) have a high chance to be promoted, but since they do it on a long time horizon they will make only one promotion in their career. In Figure 6 we report the same probability evaluated after 5 years from the start of a training period. Obviously,
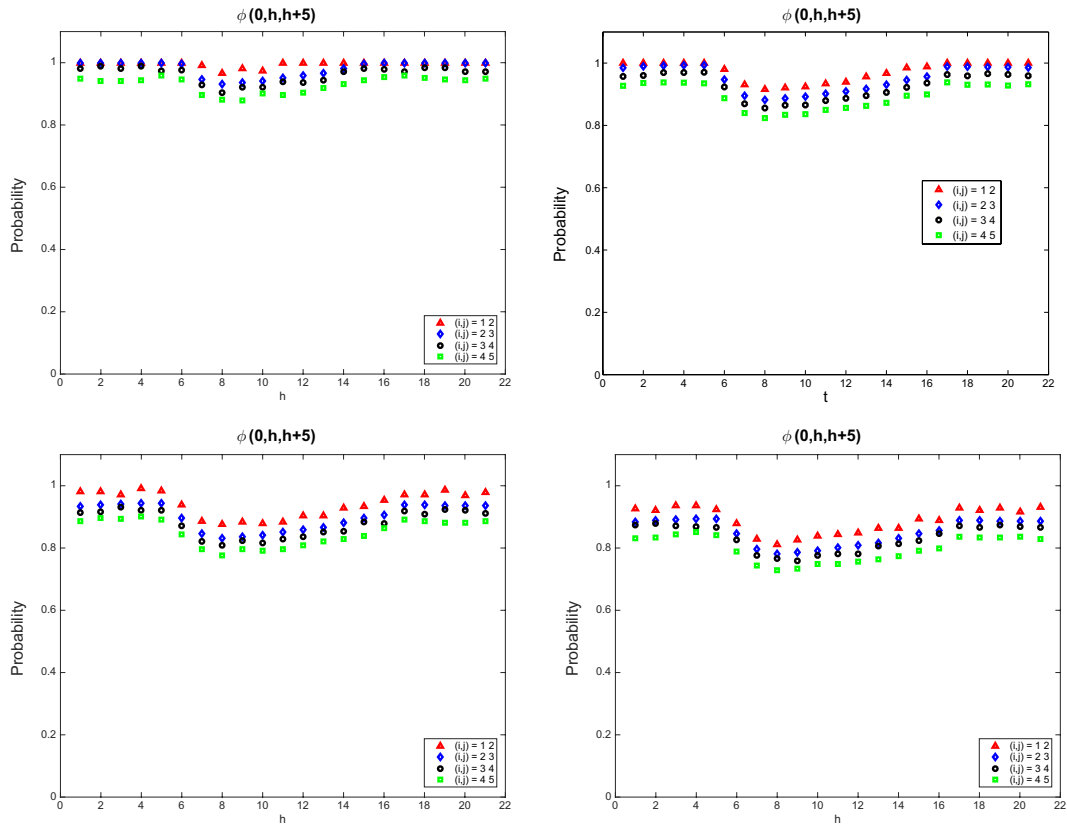
FIGURE 6. Probability function ${}^{b}\phi_{iu_1;j}(l,t)$ for $l=0$, $u_1=h$, $t=h+5$, as a function of h, for different couples of states i and j. Average shares $\frac{U}{X}$ and $\frac{V}{X}$: $70-30\%$ (*top-left*), $75-25\%$ (*top-right*), $80-20\%$ (*bottom-left*), $85-15\%$ (*bottom-right*).

the probabilities in this case are higher as compared to those in Figure 5. Also in these cases we remark a decrease of the transition probability as the percentage of time the individual is in the first step of the waiting time increases; moreover, the general shape of the functions is kept unaltered.

Finally in Figure 7 we report the probability ${}^{b}\phi_{iu_1;j}(l,t)$ for $u_1=6$, $t=10$ as a function of the backward time l for different couples of states i and j. This shows a duration effect that is important to be captured by the process. Indeed, the transition probability shows a marked dependence on the value of the backward recurrence time process.

6. CONCLUSION

In this paper, we introduced a particular type of semi-Markov models, that we called step semi-Markov models. The interest of these models comes from the fact that the sojourn times are the sum of two (or several) times. This type of stochastic tool could be particularly useful for investigating the behaviour of some particular real systems for which the sojourn times between two visited states are realised through successive "steps". We illustrate the interest of our technique through an example in manpower management.

As a continuation of the present work, several directions could be investigated.

First, in a future work we will adapt our methodology in order to investigate problems of interest in reliability and survival analysis.

Second, from a theoretical point of view, in the present paper we concentrated mainly on two-step semi-Markov models. As stated in Remark 2.6, the generalization to $m-$step semi-Markov chains can be immediately done,
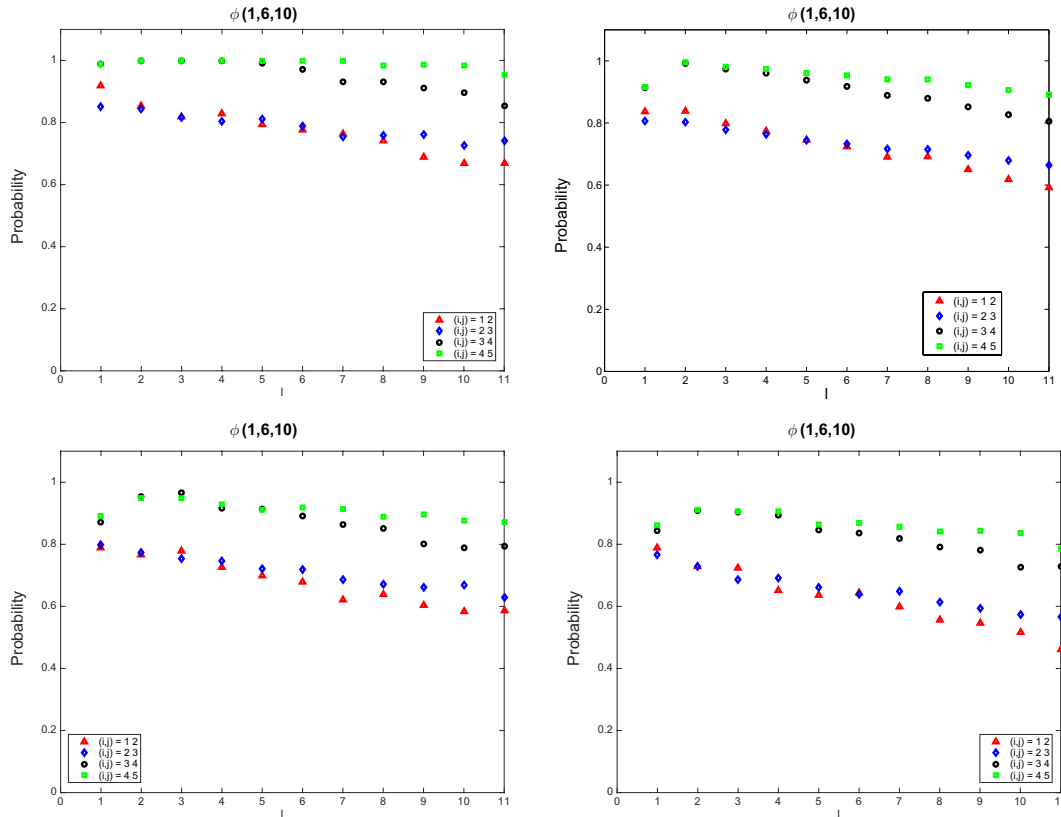
FIGURE 7. Probability function ${}^{b}\phi_{iu_1;j}(l,t)$ for $u_1 = 6$, $t = 10$, as a function of the backward time l, for different couples of states i and j. Average shares $\frac{U}{X}$ and $\frac{V}{X}$: $70-30\%$ (*top-left*), $75-25\%$ (*top-right*), $80-20\%$ (*bottom-left*), $85-15\%$ (*bottom-right*).

with the difficulty/challenge of choosing the interesting type of dependence, different numbers of steps for each transition, *etc.*

As we discussed at the end of Section 4, another important topic for further research is to undertake the estimation of the main quantities defining the model, in the case of latent variables U_n and V_n, when the only available observations of the sojourn times are $X_n = U_n + V_n, n = 1, 2, \ldots$

Other natural extensions of the models that we presented in this paper are those of non-homogeneous step semi-Markov models and continuous-time step semi-Markov processes. These represent different points that we focus on in our ongoing and future work.

Another important point would be to consider that the Bank is an open system, by taking into account what is called a Markov/semi-Markov system, that is allowing individuals to enter or leave the system. This point will be the topic of a further work.

References

[1] V.S. Barbu and N. Limnios, Semi-Markov chains and hidden semi-Markov models toward applications – their use in reliability and DNA analysis. Vol. 191 of *Lect. Notes Stat.* Springer, New York (2008).

[2] D.J. Bartholomew, *Stochastic models for social processes*. Wiley, London, 3rd edition (1982).

[3] D.J. Bartholomew, A.F. Forbes and S.I. McClean, *Statistical Techniques For Manpower Planning*. Wiley, London, 2nd edition (1991).

[4] A. Charnes, W.W. Cooper, R.J. Niehaus and D. Sholtz, Multilevel models for career management and resource planning. *Proc. of NATO Conference Manpower Planning Models*. Cambridge, September (1971).

[5] G. D'Amico, J. Janssen and R. Manca, Duration dependent semi-Markov models. *Appl. Math. Sci.* **5** (2011) 2097–2108.

[6] G. D'Amico, G. Di Biase, F. Gismondi and R. Manca, The evaluation of generalized Bernoulli processes for salary lines construction by means of continuous time generalized non-homogeneous semi-Markov processes. *Commun. Statis. – Theory and Methods* **42** (2013) 2889–2901.

[7] T. De Feyter, Modelling heterogeneity in manpower planning: dividing the personnel system into more homogeneous subgroups. *Appl. Stoch. Models Bus. Ind.* **22** (2006) 321–334.

[8] T. De Feyter and M. Guerry, Evaluating recruitment strategies using fuzzy set theory in stochastic manpower planning. *Stoch. Anal. Appl.* **27** (2009) 1148–1162.

[9] V.A. Dimitriou and N. Tsantas, Evolution of a time dependent Markov Model for training and recruitment decisions in manpower planning. *Linear Algebra and its Applications* **433** (2010) 1950–1972.

[10] V.A. Dimitriou and N. Tsantas, The augmented semi-Markov system in continuous time, *Commun. Stat.: Theory Methods* **41** (2012) 88–107.

[11] A.T. Ernst, H. Jiang, M. Krishnamoorthy and D. Sier, Staff scheduling and rostering: a review of applications, methods and models. *Eur. J. Oper. Res.* **153** (2004) 3–27.

[12] A.C. Georgiou and N. Tsantas, Modelling recruitment training in mathematical human resource planning. *Appl. Stoch. Models Bus. Ind.* **18** (2002) 53–74.

[13] F. Gismondi, R. Manca and A.V. Swishchuk, Salary lines forecasting by means of generalized binomial processes, *Int. J. Manag. Sci. Eng. Manag.* **4** (2010) 309–320.

[14] M.A. Guerry, On the evolution of stock vectors in a deterministic integer-valued Markov system. *Linear Algebra Appl.* **429** (2008) 1944–1953.

[15] J. Janssen and R. Manca, Salary cost evaluation by means of non-homogeneous semi-Markov processes. *Stoch. Models* **18** (2002) 7–23.

[16] J. Janssen and R. Manca, *Applied Semi-Markov Processes*. Springer, New York (2006).

[17] J. Janssen and R. Manca, *Semi-Markov risk models for finance, insurance and reliability*. Springer, New York (2007).

[18] J. Janssen, R. Manca and E. Volpe di Prignano, Semi-Markov modelization for salary line evolution. *Proc. of the VIII Applied Stochastic Models and Data Analysis*, edited by J. Janssen and N.C. Lauro. Anacapri (Napoli), Italy, Napoli (1997).

[19] M. Karaliopoulou, On the number of word occurrences in a semi-Markov sequence of letters. *ESAIM: PS* **13** (2009) 328–342.

[20] G.L. Lilien and A. Rao, A model for man-power management. *Manage. Sci.* **21** (1975) 1447–1457.

[21] N. Limnios and G. Oprişan, *Semi-Markov processes and reliability*. Birkhäuser, Boston (2001).

[22] S.I. McClean, A semi-Markov model for a multigrade population with Poisson recruitment. *J. Appl. Probab.* **17** (1980) 846–852.

[23] S.I. McClean, Semi-Markov models for manpower planning, In *Semi-Markov models: theory and applications*, edited by J. Janssen. Plenum (1986).

[24] S.I. McClean, Semi-Markov models for human resource modelling. *IMA J. Manage. Math.* **4** (1992) 307–315.

[25] S.I. McClean and O. Gribbin, A non-parametric competing risk model for man power planning. *Appl. Stoch. Models Data Anal.* **7** (1991) 327–341.

[26] S.I. McClean and E. Montgomery, Estimation for semi-Markov manpower models in a stochastic environment. In *Semi-Markov models and applications*, edited by J. Janssen and N. Limnios. Kluwer (2000).

[27] S.I. McClean, E. Montgomery and F. Ugwuowo, Non-homogeneous continuous-time Markov and semi-Markov manpower models. *Appl. Stoch. Models Data Anal.* **13** (1998) 191–198.

[28] A.A. Papadopolou and P.-C.G. Vassiliou, Non-homogeneous semi-Markov systems and mantainability of the state sizes. *J. Appl. Probab.* **29** (1992) 519–534.

[29] A.A. Papadopoulou and P.-C.G. Vassiliou, Asymptotic behaviour of non-homogeneous semi-Markov systems. *Linear Algebra Appl.* **210** (1994) 153–198.

[30] A.A. Papadopoulou and P.-C.G. Vassiliou, Continuous time non-homogeneous semi-Markov systems. In *Semi-Markov Models and Applications*, edited by J. Janssen and N. Limnios. Kluver Academic Publishing (1999) 241–251.