The International Journal of Biostatistics

Volume 8, *Issue* 1	2012	*Article* 23

Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi–State Models

Cristian Spitoni, *Leiden University Medical Centre*
Marion Verduijn, *Leiden University Medical Centre*
Hein Putter, *Leiden University Medical Centre*

Recommended Citation:
Spitoni, Cristian; Verduijn, Marion; and Putter, Hein (2012) "Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi–State Models," *The International Journal of Biostatistics*: Vol. 8: Iss. 1, Article 23.
DOI: 10.1515/1557-4679.1375

Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi–State Models

Cristian Spitoni, Marion Verduijn, and Hein Putter

Abstract

In this paper we discuss estimation of transition probabilities for semi–Markov multi–state models. Non–parametric and semi–parametric estimators of the transition probabilities for a large class of models (forward going models) are proposed. Large sample theory is derived using the functional delta method and the use of resampling is proposed to derive confidence bands for the transition probabilities. The last part of the paper concerns the presentation of the main ideas of the R implementation of the proposed estimators, and data from a renal replacement study are used to illustrate the behavior of the estimators proposed.

KEYWORDS: functional delta–method, semi–markov processes, survival analysis

Author Notes: Research leading to this paper was supported by the Netherlands Organization for Scientific Research Grant ZONMW-912-07-018 “Prognostic modeling and dynamic prediction for competing risks and multi–state models”. The authors are grateful to NECOSAD for providing the data. The authors are grateful to two anonymous referees and to the Associate Editor for comments and suggestions that led to substantial improvements in the paper.

Cristian Spitoni, Hein Putter: Department of Medical Statistics and Bioinformatics Leiden University Medical Centre, The Netherlands.
Marion Verduijn: Department of Clinical Epidemiology, Leiden University Medical Centre, The Netherlands.

1 Introduction

Medical investigations are often concerned with the evaluation of the effect of several time–dependent stochastic events acting on the same subject. In this direction, multi–state models describe the evolution of subjects between a finite number of states or systems exposed to several kinds of events (Andersen, 2002, Putter et al., 2007, Xu et al., 2010, Beyersmann et al., 2012). Usually multi–state models are specified in terms of transition intensities, which may involve two different time scales: the *calendar time* since the origin of the process, and the *duration time* in the current state. The corresponding models are termed *clock–forward* and *clock–reset* models, respectively, in Putter et al. (2007). The classical way to deal with *calendar time* dependence of the transition intensities, is using a Markov model. Its basic assumption is that, given the current state and the study time, the next transition is independent of the rest of the history. For such a process the transition probabilities and the transition intensities are linked by the Kolmogorov backward and forward differential equations. By solving these equations, the transition probabilities can be expressed as a function of transition intensities in the form of product integral (Aalen and Johansen, 1978, Andersen et al., 1993). However, in some applications the Markov assumption is not plausible, since transition intensities may depend on the duration time in the current state. For instance, in modeling the evolution of the human papillomavirus (HPV), which is known to be associated to cervical cancer, the Markov assumption would not account for the strong association between infection duration and progression to cervical abnormality (Kang and Lagakos, 2007). The *homogeneous semi–Markov models* are a first step toward the inclusion of the duration time dependence into the model. They assume that the future evolution of the system depends on the history only through the current state and the duration time in the current state. Many efforts have been spent to study statistical inference for these semi–Markov models. Indeed, Lagakos et al. (1978) presented nonparametric maximum likelihood estimation for the semi–Markov kernel, proposing a plug-in estimator. Their approach allows an arbitrary number of states as well as right censored observations. In Gill (1980) the theory of stochastic integration and counting processes provides a rigorous derivation of the consistency and weak convergence of the estimator of the semi–Markov kernel proposed by Lagakos et al. (1978). However, as pointed by Gill (1980), the asymptotic Gaussian process does not have an independent increment structure, thus it can not be transformed into the standard Brownian bridge or Brownian motion to construct confidence bands for the semi–Markov kernel.

Estimation of the transition probabilities for the semi–Markov homogeneous models under the Cox proportional hazards model was studied in Dabrowska (1995): more precisely, a modulated Markov renewal process was considered, whose intensities have the form of a Cox regression, with the baseline intensities being functions of the backwards recurrence time of the process. In general it was proven that in case of time fixed covariates or of time dependent covariates depending only on this reset time scale, the parameter estimates have the same asymptotic distribution as in the classical Cox regression. Moreover, in the case of time–independent covariates, thanks to the renewal nature of the process, it was possible to estimate the transition probabilities and obtain confidence intervals by bootstrapping. For the special case of the three state irreversible illness–death model, Shu et al. (2007) give explicit formulas for the transition probabilities and associated variances.

The aim of the present paper is to introduce non–parametric and semi–parametric estimators of the matrix of transition probabilities with right–censored and left–truncated data for a large class of *homogeneous* semi–Markov models. Homogeneous here refers to the assumption that transition intensities depend on duration only. Furthermore, we study the asymptotics of such estimators, providing estimates for their standard errors. The novelty of the paper is the derivation of closed form plug–in estimators of the standard errors of the transition probabilities for a large class of models (i.e. *forward–going* models). These are essentially obtained by the functional delta–method applied to certain convolution integrals of the semi–Markov kernels. For these kernels we use the estimators proposed in the papers by Gill (1980) and Dabrowska (1995) and their large sample behavior. They are the building block of our analysis: thanks to a representation theorem linking the semi–Markov kernel to the transition probabilities via a convolution functional, we are able to transfer the asymptotics of the kernels to the estimator by the application of the functional delta method. Fundamental in this regard is the evaluation of the Hadamard derivative of such convolution integrals. In order to apply our analysis to real data, we also implement in R estimates of the transition probabilities and their standard errors for a very large class of homogeneous semi–Markov models. Essentially the only limitation is that the paths of the process have to be without loops.

Finally, since for models with many states the calculation of the proposed estimator may be time consuming, we sketch the basic ideas of a resampling method (*wild bootstrap*) following Lin et al. (1994). These ideas can be used to obtain *confidence bands* rather than the pointwise confidence intervals provided by our plug–in formulas.

The paper is organized as follows. In Section 2, the basic definitions are given. In Section 3 we propose our estimator for the transition probabilities and derive their asymptotic behavior. We also give the basic ideas on how to construct confidence bands. In Section 4 we describe the implementation in R and we apply our software to data from the NECOSAD study on renal replacement therapy. Finally, Section 5 discusses further developments and how to overcome the present limitations (e.g. inclusion of *non–homogeneous* semi–Markov models and loops).

2 Notations and basic definitions

In order to define a Markov renewal process, we shall use the definitions and the formalism of Dabrowska et al. (1994) and Dabrowska (1995). Let $0 = T_0 < T_1 < T_2 < \ldots < T_m$ be consecutive times of entrance into the states $S_0, S_1, S_2, \ldots S_m \in \{1, \ldots, r\}$, then $(S,T) = (S_\ell, T_\ell)_{\ell \geq 0}$ forms a Markov renewal process if the sequence of states visited $S = (S_\ell : \ell \geq 0)$ is a Markov chain and the sojourn times $J_{m+1} = T_{m+1} - T_m$ satisfy:

$$\mathbb{P}\{S_{m+1} = j, J_{m+1} \leq \tau | S_0, T_0, \ldots, S_m, T_m\} = \mathbb{P}\{S_{m+1} = j, J_{m+1} \leq \tau | S_m\}$$

The matrix $\mathbf{Q}(\tau) := Q_{jk}(\tau) : j,k \leq r$ with

$$Q_{jk}(\tau) = \mathbb{P}\{S_{m+1} = k, J_{m+1} \leq \tau | S_m = j\}$$

is then the semi–Markov kernel of the process (S,T).

We remark that the conditional distribution of the pair $\{S_{m+1}, T_{m+1}\}$ given the whole history up to step m and the fact that the process did not end yet, depends only on $\{S_m, T_m\}$. If we denote by

$$\nu(t) := \sup\{m \geq 0 : T_m \leq t\}$$

the *counting process* registering the number of jumps made up to time t, we can finally define:

Definition 2.1 *The process $\{S(t); t \geq 0\}$ defined by*

$$S(t) := S_{\nu(t)}$$

*is called a **homogeneous semi–Markov** process determined by $\{S_\ell, T_\ell\}_\ell$.*

We now consider the semi–Markov kernel $Q_{jk}(\tau)$ and we define the following quantities useful in the sequel:

$$H_j(\tau) := \sum_{k=1}^{r} Q_{jk}(\tau) = \mathbb{P}(J_{m+1} \leq \tau | S_m = j). \tag{2.1}$$

2.1 Counting Process Formulation

Since the asymptotic results for the estimator of the semi–Markov kernel are obtained using a representation of the process in terms of *counting processes*, we give in this section the basic formalism used in order to include also right–censored data. For a given realization of the process, define $\bar{N}_{jk}(t) = \sum_{m\geq 1} I(T_m \leq t, S_m = k, S_{m-1} = j)$, the number of observed $j \to k$ transitions before time t. $\bar{N}_{jk}(t)$ is such that the sample paths are almost surely finite, right continuous integer valued with jump of size 1, with no two processes jumping at the same time and $\bar{N}_{jk}(0) = 0$. We also consider the same random censorship considered in Gill (1980) and in Dabrowska et al. (1994). Thus the times at which the process is observed are determined by the predictable zero–one process $K(t) = \sum_m I(T_m < t \leq C_m)$, where $C_m \in [T_m, T_{m+1}]$ are stopping times (e.g. censoring variables). We introduce now the counting processes which count the observed number of $j \to k$ transitions before *sojourn time u*:

$$\tilde{N}_{jk}(u) := \sharp\{m \geq 1 : S_{m-1} = j, S_m = k, T_m - T_{m-1} \leq u, K(T_m) = 1\}$$

and $\tilde{N}(u) := \sum_{j,k} \tilde{N}_{jk}(u)$. Besides, let

$$\tilde{Y}_j(u) := \sharp\{m \geq 1 : S_{m-1} = j, T_m - T_{m-1} \geq u, K(T_{m-1} + u) = 1\}$$

be the number of observed sojourn times in state j that are not smaller than u. Suppose we have n independent identically distributed observations of $\tilde{N}_{jk}(u)$ and $\tilde{Y}_j(u)$, we denote with $N_{jk}(u)$ and $Y_j(u)$ the sum over the n realizations, dropping in the notation the dependence on n. Finally, we define $N_j(u) = \sum_{k=1}^{r} N_{jk}(u)$ the total number of transitions from j.

2.2 Estimation of Semi–Markov kernel Q

The non–parametric estimation of the semi–Markov kernel $\mathbf{Q}(t)$ is given by Gill (1980): the estimator of $\widehat{\mathbf{Q}}(t)$ is derived in two steps: first estimating the survival probability $H_j(t)$ in state j by $\widehat{H}_j(t)$, defined by

$$1 - \widehat{H}_j(t) = \prod_{s \leq t} \left(1 - \frac{\Delta N_j(s)}{Y_j(s)}\right), \tag{2.2}$$

with $N_j(\cdot) := \sum_k N_{jk}(\cdot)$, and then defining $\widehat{Q}_{jk}(t)$ by

$$\widehat{Q}_{jk}(t) = \int_0^t (1 - \widehat{H}_j(s-)) \frac{dN_{jk}(s)}{Y_j(s)} \, . \tag{2.3}$$

Estimation and large sample theory for the non–parametric estimator of the kernel is reviewed in Appendix A.

We recall as well the semi–parametric estimator proposed by Dabrowska (1995), which incorporates covariates through the Cox regression form. In the present paper we restrict our attention to the case of time–independent covariates $\mathbf{Z}$ that preserve the Markov renewal properties of the process. If we define the following Breslow–like estimator for the baseline hazard:

$$\hat{A}_{jk;0}(t,\beta) := \int_0^t \frac{I(S_{jk}^{(0)}(u,\beta) > 0)}{nS_{jk}^{(0)}(u,\beta)} dN_{jk}(u)$$

with

$$S_{jk}^{(0)}(u,\beta) = n^{-1} \sum_{i=1}^{n} \tilde{Y}_j^{(i)}(u) e^{\beta^\top \mathbf{Z}_{jk}^{(i)}}, \tag{2.4}$$

where β are the regression parameters and $(\tilde{Y}_j^{(i)}, \mathbf{Z}_{jk}^{(i)})$ is the realization of the process $(\tilde{Y}_j, \mathbf{Z}_{jk})$ associated with the i–th subject. Hence, in analogy to Gill (1980), for an individual with covariate $\mathbf{Z}$, the semi–Markov kernel $\mathbf{Q}(\tau;\mathbf{Z}_{jk})$ can be written in terms of product integrals:

$$Q_{jk}(\tau;\mathbf{Z}) = \int_0^t \left(\prod_{[0,s)} (1 - dA_j(s;\mathbf{Z})) \right) dA_{jk}(s;\mathbf{Z}_{jk}), \tag{2.5}$$

where $A_j(s;\mathbf{Z}) = \sum_\ell A_{j\ell}(s;\mathbf{Z}_{j\ell})$. Under continuity conditions, (2.5) can be rewritten:

$$Q_{jk}(\tau;\mathbf{Z}) = \int_0^t \exp\{-A_j(u;\mathbf{Z})\} \, dA_{jk}(u;\mathbf{Z}_{jk}).$$

In Appendix B a more precise definition of the model is given and the large sample theory of the empirical estimator suggested by (2.5) is derived. This derivation is essentially based on the results contained in Dabrowska (1995).

3 Inference for transition probabilities

3.1 Estimation of transition probabilities

The results of this section hold both for the non–parametric and semi–parametric case. In the latter case, $\mathbf{Z}_{jk}$ will be suppressed in the notation. For homogeneous semi–Markov models we have that the transition probabilities satisfy the following Volterra integral equation of second order, also known as the Markov renewal equation:

$$P_{jk}(t) = \delta_{jk}(1 - H_j(t)) + \sum_{\ell=1}^{r} \int_0^t P_{\ell k}(t-u)\, dQ_{j\ell}(u), \tag{3.6}$$

where $P_{jk}(t) = \mathbb{P}(S(t) = k|S(0) = j)$ and $1 - H_j(t)$ is the survival probability in state j. In words, in order to find the process in state j at time t, given that it was in state i at time 0, it either should stay there (if $j = k$), or pass to any state ℓ, at some time u and take $t - u$ time units to go from there to k.

Following Çinlar (1969), if $\mathbf{M}$ is a matrix of measures and $\mathbf{V}$ a matrix of measurable functions, the convolution of $\mathbf{M}$ and $\mathbf{V}$ (written $\mathbf{M} * \mathbf{V}$) is defined by:

$$(\mathbf{M} * \mathbf{V})_{jk}(t) := \sum_{\ell} \int_0^t V_{\ell k}(t-s)\, dM_{j\ell}(s).$$

Since the matrix multiplication is not commutative, neither is the convolution. Repeated convolutions are defined in the usual manner as $\mathbf{K} * \mathbf{M} * \mathbf{V}(t) = \mathbf{K} * (\mathbf{M} * \mathbf{V})(t)$. It is straightforward that repeated convolutions between matrices are associative when all matrices involved are non-negative. Hence, the renewal equations (3.6) can be written:

$$\mathbf{P}(t) = \mathbf{h}(t) + \mathbf{Q} * \mathbf{P}(t), \tag{3.7}$$

where $\mathbf{h}(t)$ is the diagonal matrix whose diagonal elements are $h_j(t) = 1 - \sum_k Q_{jk}(t)$. In order to obtain an estimator for these transition probabilities, we want to relate them to the semi–Markov kernel in a statistically treatable form. For this reason we introduce:

$$Q_{jk}^{(m)}(\tau) = \mathbb{P}_j(S_m = k, T_m \leq \tau) = \mathbb{P}(S_m = k, T_m \leq \tau | S_0 = j). \tag{3.8}$$

$Q_{jk}^{(m)}(t)$ satisfies the following *recursive equations*, that can be written in term of convolution integrals:

$$Q_{jk}^{(m+1)}(\tau) = \sum_{\ell} \int_0^{\tau} Q_{j\ell}(du)\, Q_{\ell k}^{(m)}(\tau - u) = (\mathbf{Q} * \mathbf{Q}^{(m)})_{jk}(\tau) = (\mathbf{Q} * \ldots * \mathbf{Q})_{jk}(\tau).$$

For the solutions of the Volterra equations, a useful representation can be given in terms of the number of renewals $R_{jk}(t)$:

$$R_{jk}(\tau) = \mathbb{E}\left(\sum_m I_{(S_m = k, T_m \leq \tau)} | S_0 = j \right) = \sum_m Q_{jk}^{(m)}(\tau).$$

If we define the space $\mathscr{M}$ as the space of all function $\mathbf{V}$ bounded on finite intervals and Borel–measurable, the following theorem holds indeed (Çinlar, 1969):

Theorem 3.1 *The Markov renewal equation (3.7) has a solution $\boldsymbol{V} \in \mathscr{M}$ if and only if $\mathbf{R} * \mathbf{h} \in \mathscr{M}$. Furthermore, if $\| \sum_k Q_{j,k} \|_\infty < 1$ for some t, then $\mathbf{R} * \mathbf{h}$ exists and it is the unique solution of $\boldsymbol{V}(t) = \boldsymbol{h}(t) + \boldsymbol{Q} * \boldsymbol{V}(t)$.*

Hence, as a consequence of Theorem 3.1, the unique solutions of (3.7) can be written as:

$$\mathbf{P}(t) = (\mathbf{R} * \mathbf{h})(t). \tag{3.9}$$

The relation (3.9) links the semi–Markov kernel to the transition probabilities. Thus, it plays the role of the product integral formulation for Markov processes, where, given the cumulative hazards matrices $\mathbf{A}$ (Aalen and Johansen, 1978), the transition probabilities can be written in the form:

$$\mathbf{P}(s,t) = \prod_{(s,t]} (\mathbf{I} + \mathbf{A}(du)).$$

Hence, an estimator of the transition probability $P_{jk}(t)$ is given by

$$\widehat{P}_{jk}(t) = \sum_m \widehat{Q}_{jk}^{(m)} * (1 - \widehat{H}_k)(t). \tag{3.10}$$

At this point, the last ingredient for the statistical inference is to derive the asymptotics of the estimator (3.10) so far obtained.

3.2 Asymptotics for the transition probabilities

Since from the representation theorem, the transition probabilities we want to estimate can be viewed as a functional (precisely, an integral convolution) of the underlying process generating the data, it is natural to use the functional delta method. In this section we first recall the main results about the functional delta method, following and borrowing notation from van der Vaart (1998). We recall an important property of Hadamard differentiability: the *chain rule*. Consider the maps $\phi : \mathbb{D} \mapsto \mathbb{E}$ and $\xi : \mathbb{E} \mapsto \mathbb{F}$ that are Hadamard differentiable at θ and $\phi(\theta)$, respectively. Then the composed map $\xi \circ \phi : \mathbb{D} \mapsto \mathbb{F}$ is Hadamard differentiable at θ, and the derivative is the map obtained composing the two derivative maps. Thus, it allows to evaluate a complicated derivative, decomposing it into a sequence of basic maps, for which Hadamard differentiability is known or can be proven easily.

In our example, thanks to the representation Theorem 3.1 we can rewrite the transition probabilities in term of the semi–Markov kernels Q_{jk}. Hence, we are interested in applying the functional delta method for the map

$$\mathbf{Q} \mapsto \mathbf{P}.$$

For this reason, once we are able to calculate the Hadamard derivatives of the convolution kernels, we have the asymptotics of the transition probabilities via the delta method.

3.2.1 Hadamard derivative of $R_{jk}(t)$

Before starting with the calculation of the derivative of the expected number of renewals, we recall some elementary properties of the convolution operator. Let us introduce the linear space $\mathscr{V}_K[0,T]$, consisting of elements belonging to $\mathscr{D}^{K,K}[0,T]$, the space of $K\times K$ matrices of right–continuous and with left–hand limits processes (e.g. cadlag) in the time interval $[0,T]$, such that their variation norm is bounded. For $V_1,V_2 \in \mathscr{V}_K[0,T]$, it is easy to verify that also $V_1 * V_2(t) \in \mathscr{V}_K[0,T]$. Also the continuity (with respect to the sup norm) and the compact differentiability of the convolution operator are immediate.
Let us assume now that $\mathbf{Q} \in \mathscr{V}_T[0,T]$ and that there are maps $\varphi_1,\varphi_2 \in \mathscr{V}^K[0,T] \to \mathscr{V}^K[0,T]$. For a better understanding of the meaning of Hadamard derivative, it may be useful to try to derive formally the derivative of $\psi = \varphi_1(\mathbf{Q}) * \varphi_2(\mathbf{Q})$, component by component. Indeed, we can write:

$$\psi(\mathbf{Q})_{jk}(t) = \sum_{\ell=1}^{K} \int_0^t \varphi_2(\mathbf{Q})_{\ell k}(t-s)\, d\varphi_1(\mathbf{Q})_{j\ell}(s).$$

Thus

$$\begin{aligned}
\psi(\mathbf{Q}+\varepsilon\,\mathbf{f}_\varepsilon)_{jk}(t) &= \sum_{\ell=1}^{K} \int_0^t \varphi_2(\mathbf{Q}+\varepsilon\,\mathbf{f}_\varepsilon)_{\ell k}(t-s)\, d\varphi_1(\mathbf{Q}+\varepsilon\,\mathbf{f}_\varepsilon)_{j\ell}(s) \\
&= \sum_{\ell=1}^{K} \int_0^t \big(\varphi_2(\mathbf{Q})_{\ell k}(t-s) + \varepsilon\, \dot{\varphi}_2(\mathbf{Q})[\mathbf{f}_\varepsilon]_{\ell k}(t-s) + o(\varepsilon)\big) \\
&\cdot\quad d\big[\varphi_1(\mathbf{Q})_{j\ell} + \varepsilon\, \dot{\varphi}_1(\mathbf{Q})[\mathbf{f}_\varepsilon]_{j\ell} + o(\varepsilon)\big](s) \\
&= \psi(\mathbf{Q})_{jk}(t) \\
&+ \quad \varepsilon \left(\sum_{\ell=1}^{K} \int_0^t \dot{\varphi}_2(\mathbf{Q})[\mathbf{f}_\varepsilon]_{\ell k}(t-s)\, d\varphi_1(\mathbf{Q})_{j\ell}(s) \right.
\end{aligned}$$

$$+ \quad \sum_{\ell=1}^{K} \int_0^t \varphi_2(\mathbf{Q})_{\ell k}(t-s) d\dot{\varphi}_1(\mathbf{Q})[\mathbf{f}_\varepsilon]_{j\ell}(s) \Big) + o(\varepsilon).$$

Hence, passing formally to the limit:

$$\dot{\psi}(\mathbf{Q})[\mathbf{f}] = \lim_{\varepsilon\to 0} \frac{\psi(\mathbf{Q}+\varepsilon\,\mathbf{f}_\varepsilon) - \psi(\mathbf{Q})}{\varepsilon} = \varphi_1(\mathbf{Q}) * \dot{\varphi}_2(\mathbf{Q})[\mathbf{f}] + \dot{\varphi}_1(\mathbf{Q})[\mathbf{f}] * \varphi_2(\mathbf{Q}).$$

Actually, we can make this derivation rigorous, by proving the following result:

Proposition 3.2 *Let Q be a function with bounded variation norm and belonging to $\mathscr{V}^K[0,T]$. Moreover, let $\varphi_1, \varphi_2 \in \mathscr{V}^K[0,T] \to \mathscr{V}^K[0,T]$ and $\psi \in \mathscr{V}^K[0,T] \to \mathscr{V}^K[0,T]$ defined as the convolution $\psi(Q) = \varphi_1(Q) * \varphi_2(Q)$. If both φ_1 and φ_2 are Hadamard differentiable with derivatives $\dot{\varphi}_1$ and $\dot{\varphi}_2$ respectively, then ψ is also Hadamard differentiable, with derivative:*

$$\dot{\psi} = \dot{\varphi}_1 * \varphi_2 + \varphi_1 * \dot{\varphi}_2. \tag{3.11}$$

Proof. For any $\varepsilon > 0$, let $\mathbf{f}_\varepsilon \in \mathscr{D}^{K,K}[0,T]$ such that $\mathbf{f}_\varepsilon \to \mathbf{f}$ in $\mathscr{D}^{K,K}[0,T]$ as $\varepsilon \to 0$. Moreover, let us use the following notation:

$$\mathbf{Q}_{\varepsilon\mathbf{f}}(t) := \mathbf{Q}(t) + \varepsilon\,\mathbf{f}_\varepsilon(t).$$

Hence, we can write:

$$\begin{aligned}
\frac{1}{\varepsilon} & \quad (\psi(\mathbf{Q}_{\varepsilon\mathbf{f}}) - \psi(\mathbf{Q}))(t) \\
= & \quad \frac{1}{2}\int_0^t d\varphi_1(\mathbf{Q}(s)) \left[\frac{\varphi_2(\mathbf{Q}_{\varepsilon\mathbf{f}}) - \varphi_2(\mathbf{Q})}{\varepsilon}(t-s) - \dot{\varphi}_2(\mathbf{Q})[\mathbf{f}](t-s)) \right] \\
+ & \quad \frac{1}{2}\int_0^t \left[\frac{d\varphi_1(\mathbf{Q}_{\varepsilon\mathbf{f}}) - d\varphi_1(\mathbf{Q})}{\varepsilon}(s) - d\dot{\varphi}_1(\mathbf{Q})[\mathbf{f}](s)) \right] \varphi_2(\mathbf{Q})(t-s) \\
+ & \quad \frac{1}{2}\int_0^t \left[\frac{d\varphi_1(\mathbf{Q}_{\varepsilon\mathbf{f}}) - d\varphi_1(\mathbf{Q})}{\varepsilon}(s) - d\dot{\varphi}_1(\mathbf{Q})[\mathbf{f}](s) \right] \varphi_2(\mathbf{Q}_{\varepsilon\mathbf{f}})(t-s) \\
+ & \quad \frac{1}{2}\int_0^t d\varphi_1(\mathbf{Q}_{\varepsilon\mathbf{f}}(s)) \left[\frac{\varphi_2(\mathbf{Q}_{\varepsilon\mathbf{f}}) - \varphi_2(\mathbf{Q})}{\varepsilon}(t-s) - \dot{\varphi}_2(\mathbf{Q})[\mathbf{f}](t-s) \right] \\
+ & \quad \frac{1}{2}\int_0^t d\varphi_1(\mathbf{Q}(s)) \dot{\varphi}_2(\mathbf{Q}[\mathbf{f}])(t-s) \\
+ & \quad \frac{1}{2}\int_0^t d\dot{\varphi}_1(\mathbf{Q})[\mathbf{f}](s) \varphi_2(\mathbf{Q})(t-s) \\
+ & \quad \frac{1}{2}\int_0^t d\varphi_1(\mathbf{Q}_{\varepsilon\mathbf{f}})(s) \dot{\varphi}_2(\mathbf{Q}[\mathbf{f}])(t-s))
\end{aligned}$$

$$+ \quad \frac{1}{2}\int_0^t d\dot{\varphi}_1(\mathbf{Q})[\mathbf{f}](s)\varphi_2(\mathbf{Q}_{\varepsilon\mathbf{f}})(t-s), \tag{3.12}$$

where we have just added and subtracted terms. Hence, from the Hadamard differentiability of φ_1 and φ_2, the boundedness of the measures $d\phi_1$ and $d\phi_2$, it follows that the first four terms converge to zero as $\varepsilon \to 0$. Using the hypothesis on the measures $d\varphi$ and of the maps φ the theorem easily follows. ∎

In particular, as a consequence we have:

$$\begin{aligned} \dot{\mathbf{Q}}^{(2)} &= \mathbf{Q}*\mathbf{f}+\mathbf{f}*\mathbf{Q}, \\ \dot{\mathbf{Q}}^{(3)} &= \mathbf{Q}*\dot{\mathbf{Q}}^{(2)}[\mathbf{f}]+\mathbf{f}*\mathbf{Q}^{(2)} = \mathbf{Q}^{(2)}*\mathbf{f}+\mathbf{Q}*\mathbf{f}*\mathbf{Q}+\mathbf{f}*\mathbf{Q}^{(2)}. \end{aligned}$$

It is easy to prove by induction, using Theorem 3.2, that $\dot{\mathbf{Q}}^{(m)}(t)$ can be rewritten in the form:

$$\dot{\mathbf{Q}}^{(m)}[\mathbf{f}](t) = \sum_{\ell=1}^{m} \mathbf{Q}^{(m-\ell)}*\mathbf{f}*\mathbf{Q}^{(\ell-1)}(t). \tag{3.13}$$

In case $\mathbf{R} \in \mathscr{M}$, it makes sense to consider its Hadamard derivative. In case of forward–going models, where loops in the sample space are not allowed since each state can be visited at most once, the following lemma holds:

Lemma 3.3 *The expected number of renewals $\mathbf{R}(t)$ is Hadamard differentiable, with derivative:*

$$\dot{\mathbf{R}}[\mathbf{f}](t) = \sum_m \sum_{\ell=1}^{m} \mathbf{Q}^{(m-\ell)}*\mathbf{f}*\mathbf{Q}^{(\ell-1)}(t). \tag{3.14}$$

As pointed out in Dabrowska (1995), this property continues to hold also for non forward–going models. The difference is that the estimator will contain a number of terms increasing with the number of subjects, in such a way that the estimate is asymptotically unbiased.

In order to apply the delta method, we have to consider the Hadamard derivative at

$$\mathbf{f}(t) = n^{1/2}\left(\widehat{\mathbf{Q}}(t)-\mathbf{Q}(t)\right)$$

whose asymptotics is given in Appendix B and C for the non–parametric and semi–parametric case respectively:

$$n^{1/2}(\widehat{\mathbf{Q}}(t)-\mathbf{Q}(t)) \xrightarrow{\mathscr{D}} \mathbf{X}(t),$$

where $\mathbf{X}(t)$ is a Gaussian process. As a consequence of the delta method we also have:

$$n^{1/2}(\widehat{\mathbf{R}}(t)-\mathbf{R}(t)) \xrightarrow{\mathscr{D}} \dot{\mathbf{R}}[\mathbf{X}](t)=\sum_m \sum_{\ell=1}^{m} \mathbf{Q}^{(m-\ell)} * \mathbf{X} * \mathbf{Q}^{(\ell-1)}(t). \tag{3.15}$$

Clearly the process $\dot{\mathbf{R}}[\mathbf{X}](t)$ is Gaussian, with expectation zero as a consequence of the linearity of the expectation and from the properties of $\mathbf{X}(t)$.

3.2.2 Asymptotics for $\mathbf{P}_{jk}(t)$

In this subsection, we want to derive the large sample theory for the transition probabilities. We want to emphasize the fact that the results of this section are valid for both the non–parametric and the semi–parametric models. In fact, by the functional delta method we transfer the knowledge of the asymptotic behavior of $\mathbf{Q}$ to $\mathbf{P}$ via the differentiable map $\mathbf{Q} \rightarrow \mathbf{P}$. We derive the asymptotic variance of the asymptotic Gaussian process, in such a way that a pointwise confidence interval for the estimator may be constructed. In order to include in the same formalism for both the non–parametric and semi–parametric models, we drop the explicit dependence of the latter on the vector of covariate $\mathbf{Z}_{jk}$. Furthermore, the asymptotic variance will be written in terms of the covariance matrices of the asymptotic process $\mathbf{X}(t)$, whose expressions are contained in Appendix A for the non–parametric case and in Appendix B for the semi–parametric case. Moreover, for the sake of completeness, the estimators of the semi–Markov kernel given in Gill (1980) for the non–parametric case and in Dabrowska (1995) for the semi–parametric case are recalled and the main results reviewed in Appendix A and B respectively.

The last step for obtaining the asymptotic distribution of the transition probabilities is to apply the delta–method to equation (3.9). From Proposition 3.2, we have:

$$\dot{\mathbf{P}}(t)=\dot{\mathbf{R}} * \mathbf{h}(t)+\mathbf{R} * \dot{\mathbf{h}}(t). \tag{3.16}$$

If we denote by $\mathbf{X}(t)$ the limiting process of $n^{1/2}(\hat{\mathbf{Q}}-\mathbf{Q}) \xrightarrow{\mathscr{D}} \mathbf{X}(t)$, then by the functional delta-method and (3.16) we have the following weak convergence result:

$$n^{1/2}(\hat{\mathbf{P}}-\mathbf{P})(t) \xrightarrow{\mathscr{D}} \sum_m \sum_{\ell=1}^{m} \mathbf{Q}^{(m-\ell)} * \mathbf{X} * \mathbf{Q}^{(\ell-1)} * \mathbf{h}\,(t)-\mathbf{R} * \tilde{\mathbf{X}}\,(t) =: \mathbf{D}(t), \tag{3.17}$$

where $\tilde{\mathbf{X}}$ is the matrix with elements $\tilde{X}_{jk} := \delta_{jk} \sum_b X_{jb}$. Hence, by linearity of the convolution integrals and from the properties of stochastic integration, it follows that this limiting process $\mathbf{D}(t)$, defined in (3.17) is a zero–mean Gaussian process.

What we have to evaluate now to estimate the standard errors of the proposed estimator, is the variance for any time point t. We first denote with

$$E_{(j,k)}^{(g,h)}(s,t) := \mathbb{E}(X_{jk}(s)X_{gh}(t)), \tag{3.18}$$

the covariance matrix of the limiting process $\mathbf{X}(t)$. In order to simplify the notation in the forthcoming calculations, we give the following definitions:

$$\mathbf{D}(t) = \mathbf{D}^{(a)}(t) - \mathbf{D}^{(b)}(t), \tag{3.19}$$

where

$$\mathbf{D}^{(a)}(t) := \sum_m \sum_{\ell}^{m} \mathbf{Q}^{(m-\ell)} * \mathbf{X} * \mathbf{Q}^{(\ell-1)} * \mathbf{h}\ (t) \tag{3.20}$$

and

$$\mathbf{D}^{(b)}(t) := \mathbf{R} * \tilde{\mathbf{X}}\ (t).$$

Thus, the variance calculation can be split in the terms:

$$\operatorname{var}(D_{jk}(t)) = \operatorname{var}(D_{jk}^{(a)}(t)) + \operatorname{var}(D_{jk}^{(b)}(t)) - 2\operatorname{cov}(D_{jk}^{(a)}(t)\, D_{jk}^{(b)}(t)). \tag{3.21}$$

For the first term of (3.21) we can write:

$$\operatorname{var}(D_{jk}^{(a)}(t)) = \sum_{m_1,m_2} \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \mathbf{L}_{jk}(m_1,k_1;m_2,k_2), \tag{3.22}$$

where

$$\begin{aligned} & \mathbf{L}_{jk}(m_1,k_1;m_2,k_2) \\ = \quad & \sum_{\substack{a_1,b_1,\\ a_2,b_2}} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \\ & \times \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \int_0^t du_2 \int_0^{t-u_2} Q_{j,a_2}^{(m_2-k_2)}(ds_2) \\ & \times \int_0^{t-u_2-s_2} Q_{b_2,k}^{(k_2-1)}(dt_2) \mathbf{h}_k(t-u_1)\, \mathbf{h}_k(t-u_2)\, E_{(a_1,b_1)}^{(a_2,b_2)}(t_1,t_2)). \end{aligned} \tag{3.23}$$

As regards $\operatorname{var}(D_{jk}^{(b)}(t))$, we have:

$$\operatorname{var}(D_{jk}^{(b)}(t)) = \sum_{\ell_1,\ell_2} \int_0^t R_{jk}(ds_1) \int_0^t R_{jk}(ds_2) E_{(k,k)}^{(\ell_1,\ell_2)}(t-s_1,t-s_2), \tag{3.24}$$

For the last term, we have:

$$\mathbb{E}(D_{jk}^{(a)}(t)D_{jk}^{(b)}(t)) \quad = \quad \sum_{m_1,m_2}\sum_{k_1=1}^{m_1}\mathbf{M}_{jk}(m_1,k_1;m_2). \tag{3.25}$$

where

$$\begin{aligned}\mathbf{M}_{jk}(m_1,k_1;m_2) = & \sum_{\substack{a_1,b_1,\\ \ell}}\int_0^t du_1\int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1)\int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \\ & \times\int_0^t Q_{j,k}^{(m_2)}(ds_2)E_{(a_1,k)}^{(b_1,\ell)}(t_1,t-s_2)\,\mathbf{h}_k(t-u_1). \end{aligned}\tag{3.26}$$

Therefore, inserting (3.22)–(3.26) into (3.21), we obtain the final result for the variance of the process. The derivation of equations (3.22)–(3.26) is deferred to Appendix C.

We conclude this section with a remark. Since as we have seen during this calculation, the asymptotic process for the transition probabilities has a very complicated covariance structure, one can think of estimating the standard errors by simulation. An alternative way could be to approximate the process of interest with a simpler one, in the sense that sampling from its distribution is easier than in the original process. We give the main idea of this method in the next section.

3.3 Confidence bands via resampling

In this subsection we will construct the confidence bands via arguments similar to the ones used in Lin et al. (1994). As we have seen in the previous section, the limiting process of the semi–Markov kernel is asymptotically distributed as a zero mean Gaussian process, but with a very complicated covariance structure. Moreover, the limiting process is clearly with non– independent increments. Thus, the usual machinery for deriving confidence bands cannot be applied: the usual idea of rescaling the time in such a way that the process is equivalent to a Brownian bridge, here does not work. Hence, in order to face this problem, we approximate the cumulative hazards with a sum of normal Gaussian variables, that are equivalent to the original process thanks to the local martingale central limit theorem. Afterwards, one can apply the functional delta method and its chain rule property. We will introduce the notation for the semi–parametric case, the non–parametric case being similar. The first step is to define

$$W_{jk}(t,\beta;\mathbf{Z}_{jk}) := \Psi_{jk}(t,\beta) + (\eta_{jk}^{\star}(t,\beta))^{\top}\Sigma^{-1}U(\tau,\beta), \tag{3.27}$$

where Ψ and U are the processes of Proposition B.1. By the same proposition we have the convergence $n^{1/2}(\hat{A}_{jk}(t) - A_{jk}(t)) \to W_{jk}(t,\beta;\mathbf{Z}_{jk})$. Define the zero–mean martingales $\tilde{M}_{jk}^{(i)}(\cdot)$ associated to the counting processes $\tilde{N}_{jk}^{(i)}(\cdot)$, associated to the i–th subject with covariate vector $\mathbf{Z}_{jk}^{(i)}$:

$$\tilde{M}_{jk}^{(i)}(t) = \tilde{N}_{jk}^{(i)}(t) - \int_0^t Y_j^{(i)}(u) e^{\beta^\top \mathbf{Z}_{jk}^{(i)}} dA_{jk;0}(u). \tag{3.28}$$

By Dabrowska (1995), we can rewrite (3.27) in terms of the martingale processes, obtaining:

$$\begin{aligned} W_{jk}(t,\beta;\mathbf{Z}_{jk}) &= n^{-1/2} \sum_{i=1}^n \Bigg(\int_0^t \frac{e^{\beta^\top \mathbf{Z}_{jk}^{(i)}}}{s_{jk}^{(0)}(u,\beta)} d\tilde{M}_{jk}^{(i)}(u) \\ &+ \eta_{jk}^\star \Sigma^{-1} \sum_{l,m} \int_0^\tau (\mathbf{Z}_{jk}^{(i)} - e_{lm}(u,\beta)) d\tilde{M}_{lm}^{(i)}(u) \Bigg), \end{aligned}$$

with $s_{jk}^{(0)}$ and e_{lm} defined as in Appendix B. The idea of the approximation essentially relies on the fact that:

$$\mathbb{E}(\tilde{M}_{jk}^{(i)}(t)) = 0, \tag{3.29}$$

$$\operatorname{var}(\tilde{M}_{jk}^{(i)}(t)) = \mathbb{E}(\tilde{N}_{jk}^{(i)}(t)), \tag{3.30}$$

for any j,k and i. In fact, we basically replace $\{\tilde{M}_{jk}^{(i)}(t)\}_k$ with $\{G_{jk}^{(i)} \tilde{N}_{jk}^{(i)}(t)\}_k$, where $G_{jk}^{(i)}$ are independent standard normal variables. Any distribution for $G_{jk}^{(i)}$ will do as long as it has zero mean and unit variance (see Mammen (1992) for more details). In the bootstrap procedure, we regard $G_{jk}^{(i)}$ as random and consider all the other quantities as fixed. Hence

$$\begin{aligned} W_{jk}^*(t,\beta;\mathbf{Z}_{jk}) &= n^{-1/2} \sum_{i=1}^n \Bigg(G_{jk}^{(i)} \int_0^t \frac{e^{\beta^\top \mathbf{Z}_{jk}^{(i)}}}{s_{jk}^{(0)}(u,\beta)} d\tilde{N}_{jk}^{(i)}(u) \\ &+ \eta_{jk}^\star \Sigma^{-1} \sum_{l,m} G_{lm}^{(i)} \int_0^\tau (\mathbf{Z}_{jk}^{(i)} - e_{lm}(u,\beta)) d\tilde{N}_{lm}^{(i)}(u) \Bigg). \end{aligned} \tag{3.31}$$

The only thing one has to prove is that $W_{ij}^*(\cdot)$ and $W_{ij}(\cdot)$ converge weakly to two zero mean Gaussian processes having the same limiting distribution. We skip the details, since they are similar in spirit to the ones contained in Lin et al. (1994). Hence, in order to approximate $W_{jk}(\cdot)$, we simply obtain a large number of realizations from $W_{jk}^*(\cdot)$ by generating random samples $\{G_{jk}^{(i)}\}$, while fixing the data at their observed values.

Finally, by the functional delta–method we can derive the confidence bands for the transition probabilities, using the chain rule for the Hadamard differentiation. In fact, if $\varphi_1 : B_1 \to B_2$ and $\varphi_2 : B_2 \to B_3$ are Hadamard differentiable at $x \in B_1$ and $\varphi_1(x) \in B_2$ respectively, then $\psi = \varphi_2 \circ \varphi_1 : B_1 \to B_3$ is Hadamard differentiable at x with derivative $\dot{\varphi}_2(\varphi_1(x))[\dot{\varphi}_1(x)]$. In our case we have for the map ψ:

$$\mathbf{A} \xrightarrow{\varphi_1} \mathbf{Q} \xrightarrow{\varphi_2} \mathbf{P}.$$

We have that $\forall \mathbf{f} \in \mathscr{D}^{K,K}[0,T]$

$$[\varphi_1(\mathbf{f})]_{jk}(t) = \int_0^t e^{-\mathbf{f}_{j}(s)}\, d\mathbf{f}_{jk}(s),$$

where $\mathbf{f}_j := \sum_{k=1}^K \mathbf{f}_{jk}$. Moreover, $\forall \mathbf{g} \in \mathscr{D}^{K,K}[0,T]$

$$[\varphi_2(\mathbf{g})]_{jk}(t) = \mathbf{R}(\mathbf{g})_{jk}(t).$$

The Hadamard derivative of φ_2 is $\dot{\mathbf{R}}$, already evaluated in (3.14). As regards $\dot{\varphi}_1(\mathbf{A})$, for a test function $\mathbf{h} \in \mathscr{D}^{K,K}[0,T]$ it is given by

$$\big(\dot{\varphi}_1(\mathbf{A})[\mathbf{h}]\big)_{jk}(t) = \int_0^t e^{-\mathbf{A}_j(s)}\, d\mathbf{h}_{jk}(s) - \sum_{\ell} \int_0^t \mathbf{h}_{j\ell}(s)\, e^{-\mathbf{A}_j(s)}\, d\mathbf{A}_{jk}(s).$$

Since $n^{1/2}(\hat{\mathbf{A}} - \mathbf{A}) \xrightarrow{\mathscr{D}} \mathbf{W}$, for $\mathbf{P} = \psi(\mathbf{A})$ and $\hat{\mathbf{P}} = \psi(\hat{\mathbf{A}})$, the limiting distribution of $n^{1/2}(\hat{\mathbf{P}} - \mathbf{P})$ is given by $\dot{\psi}(A)[\mathbf{W}]$, which equals, by the chain rule, $\dot{\varphi}_2(\mathbf{Q})[\dot{\varphi}_1(\mathbf{A})[\mathbf{W}]]$. If the wild bootstrap $\mathbf{W}^*(t)$ has $\mathbf{W}(t)$ as limiting distribution, then since $\psi = \varphi_2 \circ \varphi_1$ is Hadamard differentiable, using results of Gill (1989), the limiting distribution of $n^{1/2}(\hat{\mathbf{P}} - \mathbf{P})$ is consistently estimated by $\mathbf{Z}^* = \dot{\varphi}_2(\hat{\mathbf{Q}})[\dot{\varphi}_1(\hat{\mathbf{A}})[\mathbf{W}^*]]$.

4 Application: implementation in R and data from the NECOSAD study

We apply our methods to data from the NECOSAD study, an observational renal replacement study following patients after initial dialysis (see Termorshuizen et al. (2003)), either hemodialysis (HD) or peritoneal dialysis (PD). A total of 1879 dialysis patients, 692 (37%) on PD, 1187 (63%) on HD were followed for a median of 8.7 years. Covariates considered are age (in years), categorized in three categories (age$<$ 40, 40 $\leq$age$\leq$ 60, age$>$ 60), the initial dialysis therapy (hemodialysis and peritoneal dialysis) and two covariates describing patient's comorbidity at baseline:

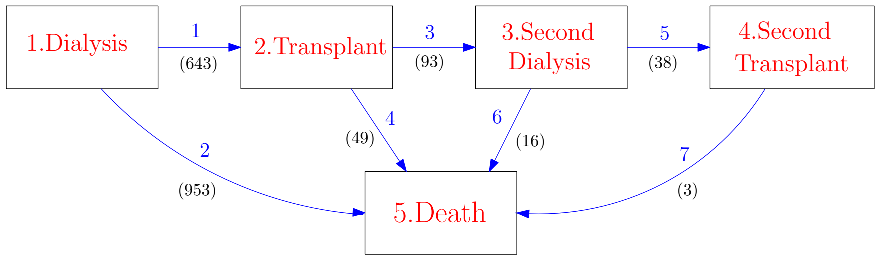

Figure 1: Multi–state model for NECOSAD data

cardiovascular disease (`cvd`) and diabetes mellitus (`dm`). The covariate frequencies in the population are shown in Table 1. The majority of the population (55%) was more than 60 years old at initiation of dialysis, one third of the population had cardiovascular disease as comorbidity, and 20% of the patients were diabetic at the start of dialysis. We model the subject's clinical history as a realization of a multi–state model, with transitions among the five states of clinical interest: dialysis (denoted with "dial" in the sequel), first kidney transplantation (Tx), relapse followed by a second dialysis, second transplantation, and death as absorbing state. The multi–state model described is shown in Figure 1, where the number of observed transitions in our data is reported in between brackets. In order to test the violation of

Table 1: Covariate frequencies in the population.

Covariate	Frequency	%
Age (yrs)		
<40	223	11.9
$40-60$	610	32.5
>60	1046	55.7
Initial dialysis therapy		
Peritoneal dialysis	692	36.8
Hemodialysis	1187	63.2
Cardiovascular disease		
No	1129	60.1
Yes	618	32.9
Diabetes mellitus		
No	1344	71.5
Yes	385	20.5

Markov renewal assumption for the multi–state model considered, one can test for each transition the significance of covariates depending on the history. We checked whether the arrival time in state 2 (the *past*) is significant for the transitions $2 \to 5$ and $2 \to 3$, by including it as a covariate in a Cox model for the transition hazards, obtaining $p = 0.12$ and $p = 0.19$ respectively. Similar results were obtained for the other transitions, so that the dataset studied can be seen as a good benchmark for our model.

Our implementation in R is aimed at extending the existing `mstate` package (for an overview see de Wreede et al. (2010)), in order to include also semi–Markov multi–state models. As we mentioned in the introduction, we want to limit our attention only to the models without loops. Hence we would like to have a criterion that, given the matrix `tmat`, tells the user whether or not the model has loops. One possible way to check this, is by defining a matrix T with $T_{jk} = 1/\#\{$ transitions from $j\}$ if $j \to k$ is possible and null otherwise, and then looking at $\exp T - \mathbf{1}$: if the diagonal terms are all null, then there are no loops. The function `Matrix.Exp` in the `msm` package (Jackson, 2011) can be used for the calculation of the exponential of a square matrix. Afterwards, one can check the maximal length p of the allowed path in the following way:

$$p = \max\{k \in \mathbb{N} : T^k \neq \mathbf{0}\}.$$

For instance, in our dataset the maximal path length p is 4, attained for the path $1 \to 2 \to 3 \to 4 \to 5$.

In case of non–parametric models, following Gill (1980), the semi–Markov kernel $Q_{jk}(t)$ can be estimated in two steps: at first we estimate the survival probabilities in the state j at time t with the estimator $1 - \widehat{H}_j(t)$ given in (A-1), then by (A-2) we obtain $\widehat{Q}_{jk}(t)$. In (A-1) we estimate the hazard for the transition $j \to k$ with the non–parametric estimator:

$$d\widehat{\mathbf{A}}_{jk}(t) = \frac{dN_{jk}(t)}{Y_j(t)}.$$

Summing over all event times up to time t, one obtains the Nelson–Aalen estimator $\widehat{\mathbf{A}}_{jk}(t)$ of the cumulative hazard for the transition $j \to k$. Estimates for the cumulative hazards (with and without covariates) can be obtained using the `mstate` function `msfit`: after having fitted a model stratified on the transitions by means of the function `coxph` from the `survival` package, this fitted model is the input of `msfit`, a function calculating transition hazards $\widehat{\mathbf{A}}_{jk}(t)$ and associated standard errors at all event points. The plot of the non–parametric cumulative transition hazards are shown in Figure 2.

We can now proceed with the implementation of the estimator for the transition probabilities. Since the maximal length of an allowed path is 4, the sum (3.10) consists of a finite number of terms:

$$\widehat{\mathbf{P}}_{jk}(t) = \sum_{m=1}^{4} \widehat{Q}_{jk}^{(m)} * (1 - \widehat{H}_k)(t).$$

Hence, one has just to estimate convolution integrals, with the same convolving kernel Q. For example, the convolved term $Q_{jk}^{(2)}(t)$ can be estimated by

$$\widehat{Q}_{jk}^{(2)}(t) = \sum_{s=1}^{5} \sum_{\ell} \Delta \widehat{\mathbf{A}}_{js}(t_\ell)(1 - \hat{H}_j(t_\ell))\hat{Q}_{sk}(t - t_\ell).$$

Figure 3 shows a plot of the estimated death probabilities $\hat{P}_{15}(t)$ and $\hat{P}_{25}(t)$, along

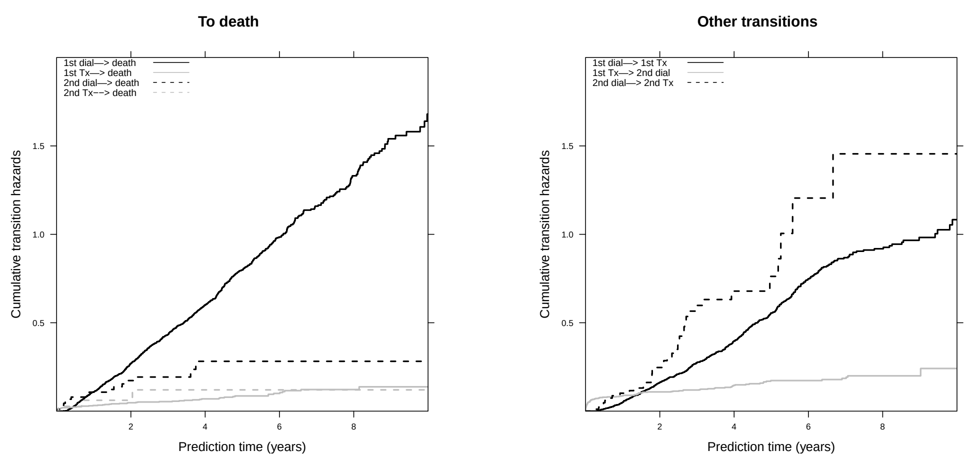

Figure 2: Non–parametric cumulative transition hazards

with 95% pointwise confidence intervals, obtained from the asymptotic results of Section 3.2. The pointwise confidence intervals were obtained on the log–scale, for instance the upper curve of the confidence interval of $\hat{P}_{15}(t)$ was obtained by $\exp\left(\log \hat{P}_{15}(t) + 1.96 \times \mathrm{SE}(\log \hat{P}_{15}(t))\right)$, with $\mathrm{SE}(\log \hat{P}_{15}(t)) = \mathrm{SE}(\hat{P}_{15}(t))/\hat{P}_{15}(t)$ given by the delta–method. As we have seen, once we have the semi–Markov kernels and the convolved kernels, it is straightforward to estimate the transition probabilities. Hence, for the non–parametric estimator of the transition probabilities, we can summarize our findings in Figure 4, which shows stacked probability plots of $\hat{P}_{1k}(t)$ (left) and $\hat{P}_{2k}(t)$ (right). The distance between two consecutive curves represents the probabilities. The most striking difference between the left and right panel is the much larger death probability from 1^{st} dialysis ($\widehat{P}_{15}(t)$) compared to

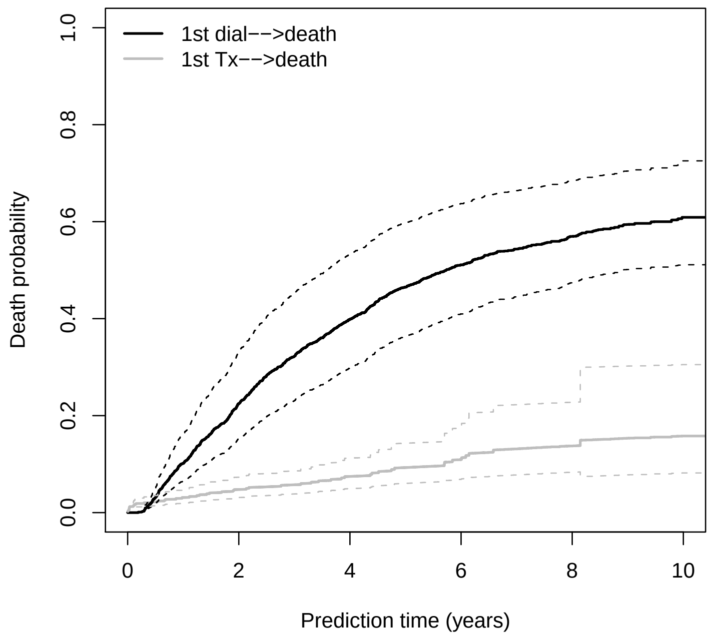

Figure 3: Non–parametric death probabilities with 95% confidence interval

the death probability from 1^{st} Tx ($\widehat{P}_{25}(t)$). One should be careful in interpreting this causally as a positive treatment effect of transplantation, since only relatively healthy patients are considered for transplantation.

As regards the semi–parametric estimator, we allowed the effects of covariates to be different across transitions. Given the low number of observed transitions, we did not include covariate effects for the transitions 2^{nd} dialysis$\rightarrow$ death and 2^{nd} Tx$\rightarrow$ death. We note that none of the covariate effects were significant for these transitions and they were subsequently removed. The effect of the covariates on the transitions are summarized in Table 2. The table confirms some known clinical facts in renal transplantation; it clearly reflects that both in patients starting their trajectory on renal replacement therapy on hemodialysis as well as in patients starting on peritoneal dialysis, increased age and the presence of comorbid conditions are contra–indications for (being put on a waiting list of) renal transplantation, while being clear risk factors for death. No clear effect of these factors were found for restart of dialysis after having had a renal transplant. This can be explained by the fact that transplant rejection is a main underlying reason, with histoincompatibility known as important risk factor.

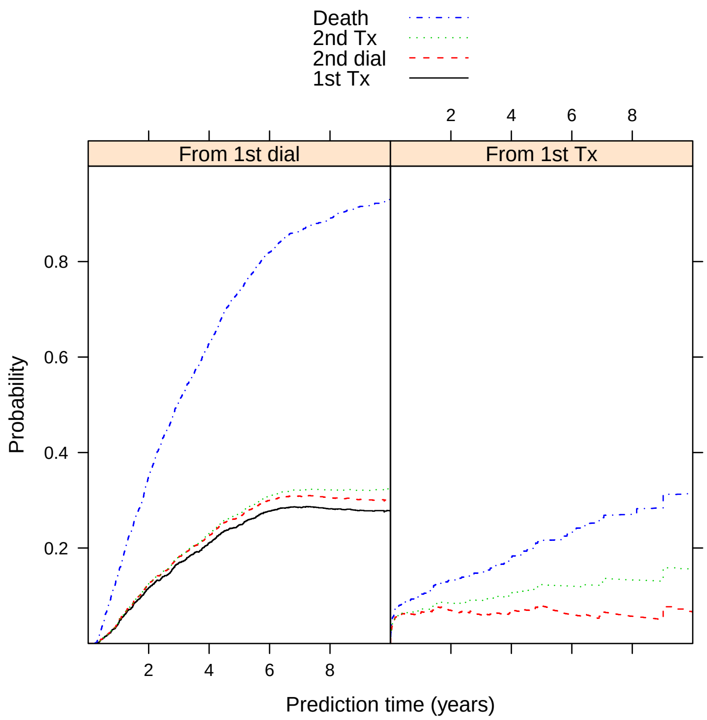

Figure 4: Stacked non–parametric transition probabilities

Table 2: Effect of the covariates on each transition.

	Category	1^{st}dial → 1^{st}Tx			1^{st}dial → death			1^{st}Tx → 2^{nd}dial			1^{st}Tx → death			2^{nd}dial → 2^{nd}Tx		
		B	(SE)	p	B	(SE)	p	B	(SE)	p	B	(SE)	p	B	(SE)	p
	Age $<$ 40															
	Age 40 – 60	-0.341	(0.154)	0.027	1.901	(0.587)	0.001	-0.217	(0.382)	0.57	1.493	(0.745)	0.04	1.049	(0.658)	0.11
HD	Age $>$ 60	-1.854	(0.183)	$<$0.0001	2.710	(0.580)	$<$0.0001	0.034	(0.446)	0.94	1.922	(0.771)	0.012	-0.459	(0.977)	0.64
	CV disease	-0.495	(0.168)	0.003	0.436	(0.085)	$<$0.0001	0.387	(0.416)	0.35	0.564	(0.381)	0.14	0.501	(0.601)	0.41
	Diabetes	-0.659	(0.202)	0.002	0.442	(0.085)	$<$0.0001	-0.829	(0.732)	0.26	0.397	(0.459)	0.39	-12.0	(70.5)	0.99
	Age $<$ 40															
	Age 40 – 60	-0.163	(0.123)	0.18	0.916	(0.312)	0.003	0.159	(0.326)	0.62	0.693	(1.125)	0.54	-0.743	(0.596)	0.21
PD	Age $>$ 60	-1.583	(0.203)	$<$0.0001	1.708	(0.305)	$<$0.0001	0.012	(0.564)	0.98	2.50	(1.13)	0.027	1.400	(0.797)	0.079
	CV disease	-0.583	(0.200)	0.003	0.658	(0.138)	$<$0.0001	-0.098	(0.539)	0.85	2.234	(0.664)	0.0008	0.206	(0.801)	0.80
	Diabetes	-0.105	(0.180)	0.56	0.605	(0.143)	$<$0.0001	0.381	(0.419)	0.36	0.287	(0.811)	0.72	-0.93	(1.04)	0.37

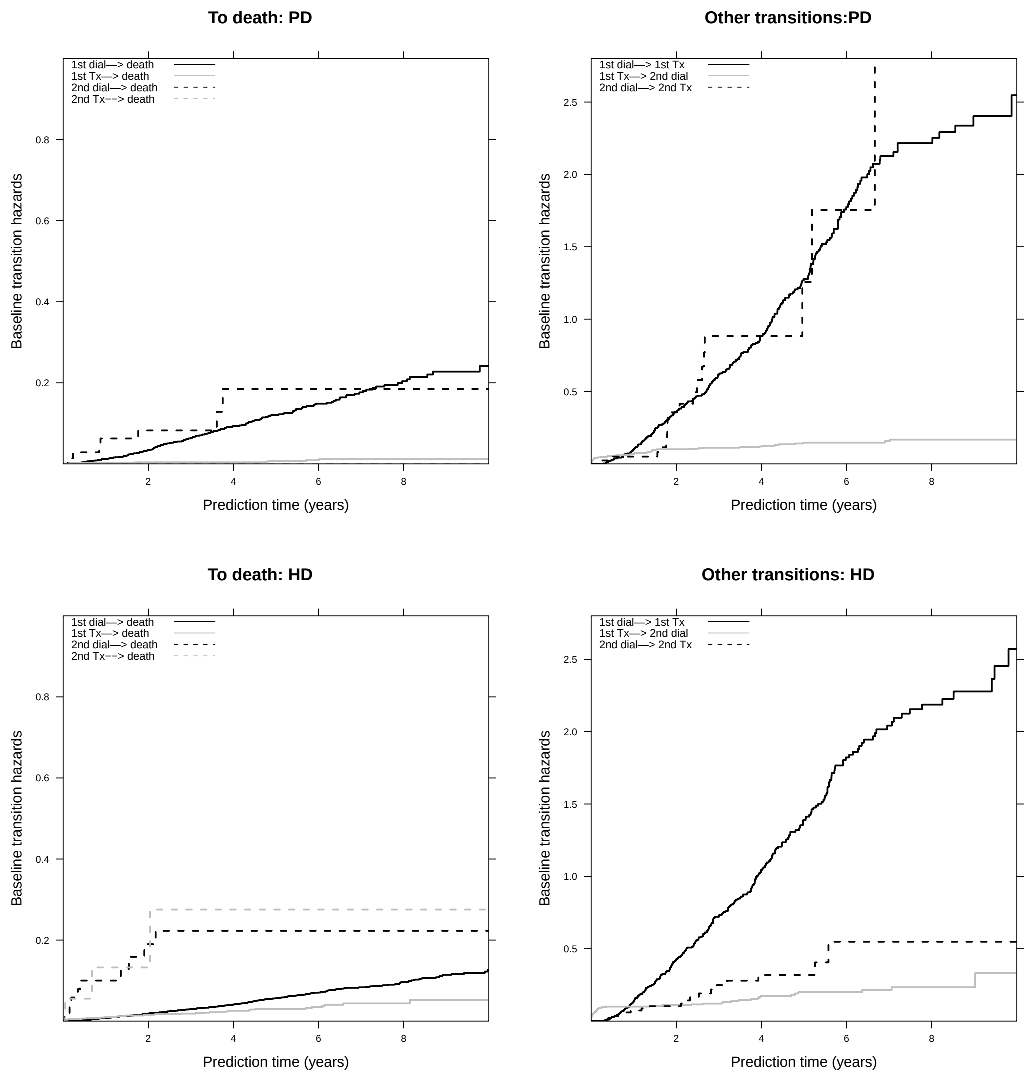

Figure 5: Cumulative baseline transition hazards

Figure 5 shows the estimated cumulative baseline transition hazards for this model. To illustrate prediction in this model, we consider three patients: A with age< 40, B with age $40-60$ and C with age > 60, all without diabetes and without cardiovascular disease. In Figure 6 we show the transition probabilities for subjects starting from first dialysis under peritoneal dialysis (PD) and hemodialysis (HD) and in Figure 7 the transition probabilities considered are from the first transplant.

The larger death probabilities starting from 1^{st} dialysis compared to starting from 1^{st} Tx, shown in Figure 4, are also present here. Also in both Figure 6 and Figure 7 we see increasing death probabilities with increasing age, in accordance with the results of Table 2.

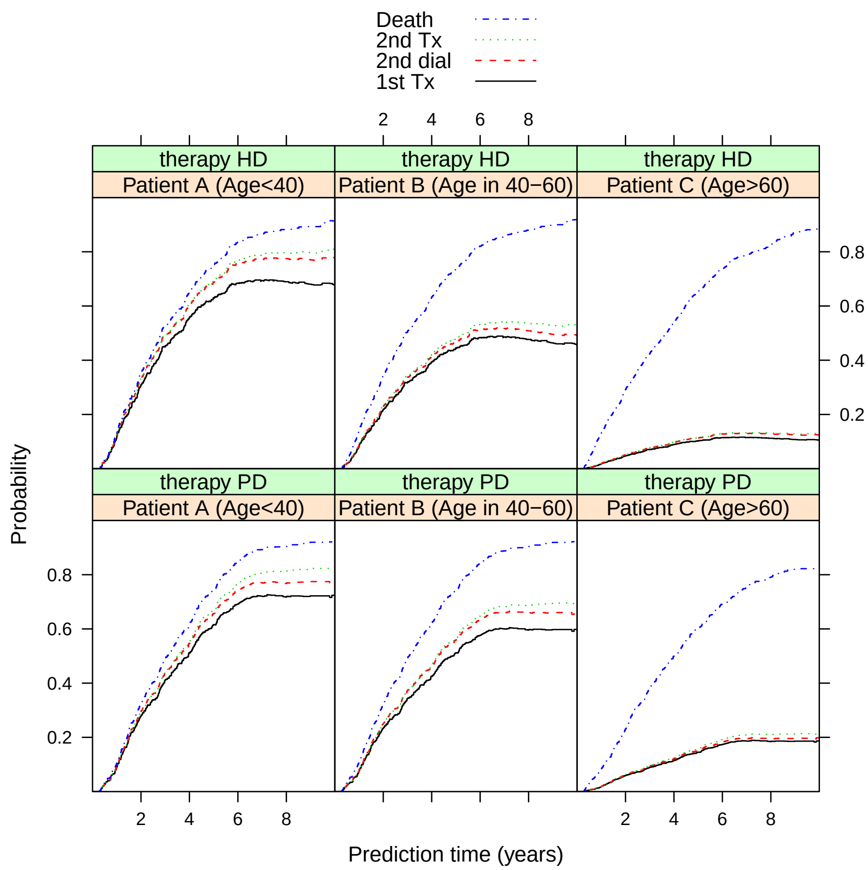

Figure 6: Stacked transition probabilities from first dialysis

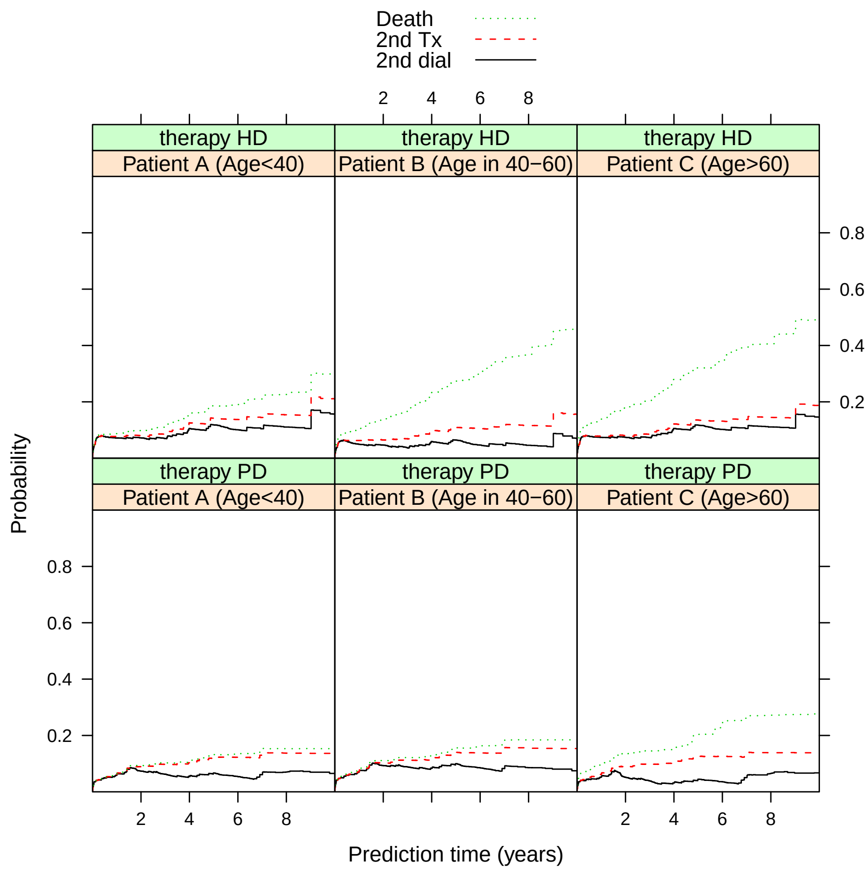

Figure 7: Stacked transition probabilities from first transplant

Finally, we want to illustrate the application of the resampling technique to derive pointwise confidence intervals in the presence of covariates. If we consider both dialysis states (states 1, first dialysis, and 3, second dialysis, in Figure 1) to be equally (un)desirable, interest could be in the estimation of the probability of being in dialysis. Note that this idea is similar to the popular current leukemia free–survival in bone marrow transplantation (Craddock et al., 2000). For a patient starting in 1^{st} dialysis, we would be interested in $P_{11}(t)+P_{13}(t)$, while for a patient starting in 2^{nd} dialysis, this probability would be $P_{33}(t)$. We sample independent standard normal variables $G_{jk}^{(i)}$ per each transition, per each individual and we estimate $\mathbf{W}^*(t)$ in (3.31). Subsequently, $\mathbf{X}^*(t)$ and $\mathbf{Z}^*(t)$ are calculated. Since $\mathbf{Z}^*(t)$ serves as an approximation to the distribution of $n^{1/2}(\widehat{\mathbf{P}}(t)-\mathbf{P}(t))$, the $\alpha/2$ and $1-\alpha/2$ quantiles, $\mathbf{z}^*_{\alpha/2}(t)$ and $\mathbf{z}^*_{1-\alpha/2}(t)$, of the bootstrap replications of $\mathbf{Z}^*(t)$ may be used as approximations of those of $n^{1/2}(\widehat{\mathbf{P}}(t)-\mathbf{P}(t))$, leading to the $(1-\alpha)100\%$

bootstrap confidence interval: $\left(\widehat{\mathbf{P}}(t) - \mathbf{z}^*_{1-\alpha/2}(t)/\sqrt{n},\ \widehat{\mathbf{P}}(t) - \mathbf{z}^*_{\alpha/2}(t)/\sqrt{n}\right)$. Figure 8 shows the results based on $B = 100$ replications, for a patient under hemodialysis with age below 40 years and no comorbidities. In Figure 8 we also reported the confidence intervals obtained by the asymptotic distribution of the transition probabilities. Note that the bootstrap confidence intervals of the bootstrap are wider, probably due to a small sample effect. A small remark about these two different approaches. Despite the complexity of the limiting process, the R implementation of the covariance matrix is straightforward and the computation time is reasonable. As regards the bootstrap approach, it allows to derive approximated confidence bands, results essentially out of reach with a plug–in estimator. On the other hand, if one is interested in the computation of pointwise confidence intervals, the wild bootstrap method is time–consuming, mostly due to the multidimensional convolution integrals that one has to evaluate. Hence, in this case the use of the plug–in estimator for the variance of transition probabilities is recommended.

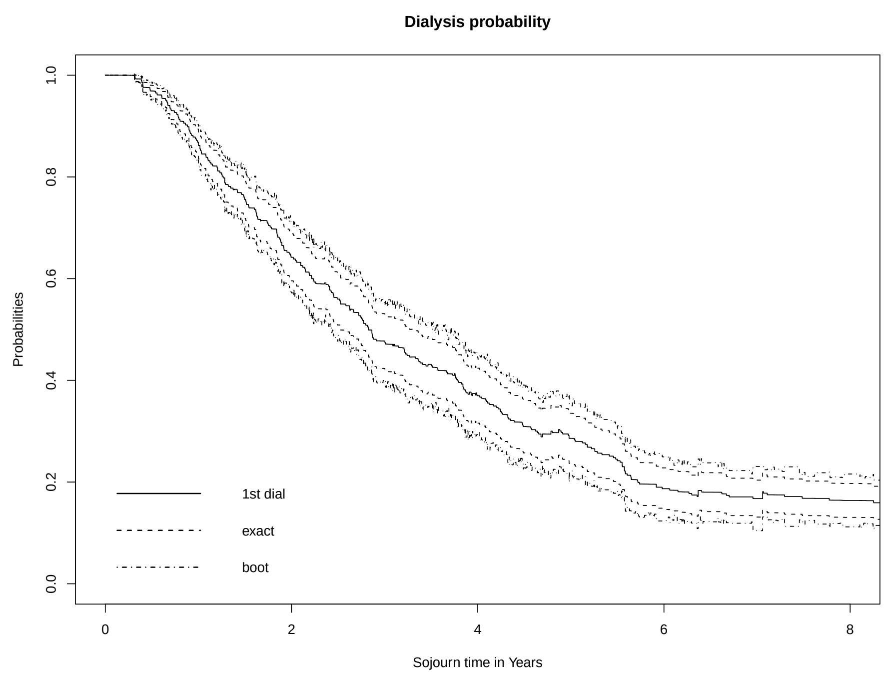

Figure 8: Probability of being in dialysis with 95% confidence intervals calculated by asymptotic results and via wild bootstrap

5 Discussion

A first natural generalization of the present paper would be the extension to *non–homogeneous* semi–Markov models. In Monteiro et al. (2006) the authors estimate the transition matrices for this class of models for window–censored data based on non–parametric estimators of the cumulative transition hazards. However, besides the limitation on the censoring mechanism, they do not derive the confidence intervals. We think that the theoretical results of this paper can be generalized also to the non–homogeneous case. The main difficulty is the implementation, since our approach of evaluating efficiently the convolution integral, is strictly related to the univariate nature of the time.

A further step in generalizing our analysis regards the inclusion of loops in the software developed. As we have already mentioned before, the limitation on the allowed paths comes only from the implementation. From a theoretical point of view indeed, the results from Gill (1980) and Dabrowska (1995) hold also in the case of loops. The inclusion of paths with loops may be possible, provided that one gives some *a priori* criterium to stop the sum of the convoluted kernels after a finite number of iterations. This can be done fixing a value ε, and determine the maximal path length in a such way that

$$m_{max} := \min\{m : \sup_t \hat{Q}^m_{jk}(t) < \varepsilon\}$$

$\forall j,k$. Of course this cut–off will be used only at the implementation level. All the estimates will remain an infinite sum and the model will not become a hierarchical model, where each possible sequence of states visited has the property that none of the states can be reached from the subsequent states.

A Estimation and large sample theory for $\mathbf{Q}(t)$ (non–parametric)

A.1 Estimation of semi–Markov kernel

In this section we review the main results from Gill (1980) about statistical inference for the semi–Markov kernel. Since for the semi–Markov kernel we can write:

$$\begin{aligned} Q_{jk}(t) &= \mathbb{P}(S_{m+1} = k, J_{m+1} \leq t | S_m = j) \\ &= \int_0^t \mathbb{P}(S_{m+1} = k, J_{m+1} \in (u, u + du] | S_m = j) \end{aligned}$$

$$= \int_0^t \mathbb{P}(S_{m+1} = k, J_{m+1} \in (u, u+du] | S_m = j, J_{m+1} > u) \, \mathbb{P}(J_{m+1} > u | S_m = j)$$

with $J_{m+1} := T_{m+1} - T_m$, the estimators of Q_{jk} is built up in two steps: first estimating H_j by $\widehat{H}_j$, defined by

$$1 - \widehat{H}_j(t) = \prod_{s \le t} \left(1 - \frac{\Delta N_j(s)}{Y_j(s)}\right) \qquad \text{(A-1)}$$

with $N_j(\cdot) := \sum_k N_{jk}(\cdot)$ and afterwards we can define $\widehat{Q}_{jk}(t)$ by

$$\widehat{Q}_{jk}(t) = \int_0^t (1 - \widehat{H}_j(s-)) \frac{dN_{jk}(s)}{Y_j(s)} \ . \qquad \text{(A-2)}$$

For a real valued stochastic process $X = \{X(t); t \in [0, \infty)\}$ whose sample path have left hand limits, X_- is the process defined by $X_-(0) = 0$ and $X_-(t) = X(t-)$. In dealing with an indexed family, we write X_{i-} for $(X_i)_-$. Moreover, ΔX is the process $X - X_-$. For the estimator $\widehat{Q}_{jk}(t)$ we have the following weak convergence theorem:

Theorem A.1 *Suppose that the number of observed sojourn times is almost surely finite and choose $\tau_j, j \le r$ such that $\mathbb{E}Y_j(\tau_j) > 0$. Then considered as a random element of $\prod_j (D[0, \tau_j])^{r+1}$, $\{n^{1/2}(\widehat{Q}_{jk}(t) - Q_{jk}(t), \, n^{1/2}(\widehat{H}_j(t) - H_j(t)\}$ is asymptotically distributed as*

$$\Big\{ \int_{(0,t]} \frac{1 - H_{j-}(s)}{\mathbb{E}Y_j(s)} dW_{jk}(s) - Q_{jk}(t) \int_{(0,t)} \frac{1 - H_{j-}(s)}{1 - H_j(s)} \frac{1}{\mathbb{E}Y_j(s)} dW_j(s)$$
$$+ \int_{(0,t)} Q_{jk}(s) \frac{1 - H_{j-}(s)}{1 - H_j(s)} \frac{dW_j(s)}{\mathbb{E}Y_j(s)}, (1 - H_j(t)) \int_{(0,t]} \frac{1 - H_{j-}(s)}{1 - H_j(s)} \frac{dW_j(s)}{\mathbb{E}Y_j(s)} \Big\},$$

where the W_{jk} are jointly zero mean Gaussian processes with independent multivariate increments, the sets $\{W_{jk}; k \le r\}$, $j = 1, \ldots, r$ being independent of one another; $W_j = \sum_k W_{ik}$; and

$$\mathrm{var}(W_{jk}(t)) = \int_0^t \mathbb{E}Y_j(s) \left(1 - \frac{\Delta Q_{jk}(s)}{1 - H_{j-}(s)}\right) \frac{dQ_{jk}(s)}{1 - H_{j-}(s)}, \qquad \text{(A-3)}$$

$$\mathrm{cov}(W_{jk}(t), W_{jk'}(t)) = -\int_0^t \mathbb{E}Y_j(s) \frac{\Delta Q_{jk}(s)}{1 - H_{j-}(s)} \frac{dQ_{jk'}(s)}{1 - H_j(s-)}. \qquad \text{(A-4)}$$

The integral with respect to W_{jk} and W_j are stochastic integrals in the sense of Meyer (the W_{jk} are square integrable martingales with respect to the natural family of σ–algebras) or can equivalently be defined by formal integration by parts.

A.2 Asymptotic covariance of semi–Markov kernel

We define:

$$D\mathbf{Q} := n^{1/2}(\widehat{\mathbf{Q}} - \mathbf{Q})(t) \tag{A-5}$$

we recall that from Theorem A.1 $D\mathbf{Q}(t)$ is asymptotically distributed as:

$$D\mathbf{Q}(t) \xrightarrow{\mathscr{D}} \mathbf{X}(t),$$

with

$$X_{jk}(t) := \int_0^t a_j(s)\, dW_{jk}(s) - \int_0^{t-} b_{j,k}(s)\, dW_j(s), \tag{A-6}$$

where

$$a_j(s) := \frac{1 - H_j(s-)}{\mathbb{E}Y_j(s)} \tag{A-7}$$

and

$$b_{j,k}(s) := \frac{Q_{jk}(t) - Q_{jk}(s)}{\mathbb{E}Y_j(s)} \frac{1 - H_j(s-)}{1 - H_j(s)}, \tag{A-8}$$

where W_{jk} are jointly zero mean Gaussian processes with independent multivariate increments. Moreover:

$$\begin{aligned} \mathrm{var}(W_{jk}(s)) &= \int_0^s dc_{j,k}(u), && \text{(A-9)} \\ \mathrm{cov}(W_{jk}(s)W_{j,k'}(s)) &= \int_0^s dd_j^{k,k'}(u)\ k \neq k', && \text{(A-10)} \\ \mathrm{cov}(W_{jk}(s)W_{j',k'}(s')) &= 0 \ \ j \neq j' \text{ or } \ s \neq s', && \text{(A-11)} \end{aligned}$$

where we define

$$dc_{j,k}(s) := \mathbb{E}Y_j(s)\left(1 - \frac{\Delta Q_{jk}(s)}{1 - H_j(s-)}\right) \frac{dQ_{jk}(s)}{1 - H_j(s-)} \tag{A-12}$$

and

$$dd_j^{k,k'}(s) := -\mathbb{E}Y_j(s) \frac{\Delta Q_{jk}(s)}{1 - H_j(s-)} \frac{dQ_{jk'}(s)}{1 - H_j(s-)}. \tag{A-13}$$

The goal of this Appendix is to evaluate the following expectation:

$$E_{(j,k)}^{(l,m)}(z,w) := \mathbb{E}(X_{jk}(z) X_{lm}(w)). \tag{A-14}$$

Thus, using these new notation for the theorem, we have:

$$E_{(j,k)}^{(l,m)}(z,w) = \mathbb{E}\Bigg[\left(\int_0^z a_j(s)\, dW_{jk}(s) - \int_0^z b_{j,k}(s)\, dW_j(s)\right)$$

$$\begin{aligned} &\times \quad \Big(\int_0^w a_l(t)\, dQ_{l,m}(t) - \int_0^w b_{l,m}(t)\, dW_l(t) \Big) \Big] \\ &=: \quad \mathscr{A}_{j,k}^{l,m}(z,w) - \mathscr{B}_{j,k}^{l,m}(z,w) - \mathscr{C}_{j,k}^{l,m}(z,w) + \mathscr{D}_{j,k}^{l,m}(z,w). \quad \text{(A-15)} \end{aligned}$$

Thus, thanks to the properties of the process $W_{j,k}$, we can easily evaluate each term of (A-15):

$$\begin{aligned} \mathscr{A}_{j,k}^{l,m}(z,w) &:= \mathbb{E}\Big(\int_0^z \int_0^w a_j(s)\, a_l(t)\, dW_{j,k}(s)\, dW_{l,m}(t) \Big) \\ &= \delta_{j,l} \int_0^{z\wedge w} a_j^2(s) \, \Big((1-\delta_{km}) dd_j^{k,m}(s) + \delta_{km} dc_{j,k}(s) \Big) \quad \text{(A-16)} \end{aligned}$$

$$\begin{aligned} \mathscr{B}_{j,k}^{l,m}(z,w) &:= \mathbb{E}\Big(\int_0^z \int_0^w a_j(s)\, b_{l,m}(t)\, dW_{jk}(s)\, dW_l(t) \Big) \\ &= \delta_{j,l} \sum_\ell \int_0^{z\wedge w} a_j(s)\, b_{j,m}(s) \, \Big((1-\delta_{k\ell}) dd_i^{k,\ell}(s) + \delta_{k\ell} dc_{j,k}(s) \Big) \quad \text{(A-17)} \end{aligned}$$

$$\begin{aligned} \mathscr{C}_{j,k}^{l,m}(z,w) &:= \mathbb{E}\Big(\int_0^z \int_0^w a_l(s)\, b_{j,k}(t)\, dW_j(s)\, dW_{l,m}(t) \Big) \\ &= \delta_{j,l} \sum_\ell \int_0^{z\wedge w} a_l(s)\, b_{j,k}(s) \, \Big((1-\delta_{\ell m}) dd_j^{\ell,m}(s) + \delta_{\ell m} dc_{j,m}(s) \Big) \quad \text{(A-18)} \end{aligned}$$

$$\begin{aligned} \mathscr{D}_{j,k}^{l,m}(z,w) &= \mathbb{E}\Big(\int_0^z \int_0^w b_{j,k}(s)\, b_{l,m}(t)\, dW_j(s)\, dW_l(t) \Big) \\ &= \delta_{j,l} \sum_{\ell,r} \int_0^{z\wedge w} b_{j,k}(s)\, b_{j,m}(s) \, \Big((1-\delta_{\ell r}) dd_j^{\ell,r}(s) + \delta_{\ell r} dc_{j,r}(s) \Big) \quad \text{(A-19)} \end{aligned}$$

Hence, inserting (A-16), (A-17), (A-18), (A-19) in (A-15), we have the expression of the correlations $E_{(i,j)}^{(l,m)}(z,w)$ in terms of the basic quantities of the limiting process $\mathbf{Z}(t)$:

$$E_{(j,k)}^{(l,m)}(z,w) = \mathscr{A}_{j,k}^{l,m}(z,w) - \mathscr{B}_{j,k}^{l,m}(z,w) - \mathscr{C}_{j,k}^{l,m}(z,w) + \mathscr{D}_{j,k}^{l,m}(z,w). \quad \text{(A-20)}$$

B Estimation and large sample theory for $\mathbf{Q}(t)$ (semi–parametric)

B.1 Estimation of semi–Markov kernel

In this subsection we recall the main results of Dabrowska (1995) about semi–parametric semi–Markov models which incorporate covariates through the Cox regression form. We restrict our attention to the case of time–independent covariates that preserve the Markov renewal properties of the process. We remark that, despite the fact that in the more general setting of time–dependent covariates dependent on the backwards recurrence time (Dabrowska, 1995) large sample theory for the cumulative hazards is known, the estimation of the transition probabilities in this case is out of reach.

Let $S(t)$ be the state occupied by the process at time t and let $L(t)$ be the backward recurrence time, in other words the time elapsed between t and the last jump of the process before t. Thus, under the Cox proportional hazards model assumptions, the evolution of the process $\tilde{N}_{jk}(\cdot)$ is determined by the intensities

$$d\mathbf{A}_{jk}(t;\mathbf{Z}_{jk}) = I(S(t-)=j)\, e^{\beta^\top \mathbf{Z}_{jk}}\, \alpha_{jk;0}(L(t)), \tag{B-1}$$

where $\alpha_{jk;0}(L(t))$ is a baseline hazard function and β a vector of coefficients associated to the vector $\mathbf{Z}$. In Dabrowska et al. (1994) it was shown that the estimates of the regression parameters β and the baseline cumulative hazards $A_{0;jk}(t) = \int_0^t \alpha_{jk;0}(u)\,du$ can be derived by profile likelihood in the same way as in the ordinary Cox regression (Andersen et al., 1993). Indeed if we define the following Breslow–like estimator for the baseline hazard:

$$\hat{A}_{jk;0}(t,\beta) := \int_0^t \frac{I(S^{(0)}_{jk}(u,\beta)>0)}{n S^{(0)}_{jk}(u,\beta)} dN_{jk}(u)$$

with

$$S^{(0)}_{jk}(u,\beta) = n^{-1}\sum_{i=1}^{n} \tilde{Y}^{(i)}_j(u) e^{\beta^\top \mathbf{Z}^{(i)}_{jk}}, \tag{B-2}$$

where $(\tilde{Y}^{(i)}_j, \mathbf{Z}^{(i)}_{jk})$ is the realization of the process $(\tilde{Y}_j, \mathbf{Z}_{jk})$ associated with the i–th subject, the following theorem has been proved in Dabrowska (1995) (Proposition 2.1) to which we refer the reader for the proof and details:

Proposition B.1 *$n^{1/2}(\hat{\beta}-\beta)$ is asymptotically distributed as $\Sigma^{-1}(\tau,\beta)U(\tau,\beta)$ and $n^{1/2}(\hat{A}_{jk;0}(t,\hat{\beta}) - A_{jk;0}(t,\beta))$ as $\Psi_{jk}(t,\beta) + \eta^\top_{jk}(t,\beta)\Sigma^{-1}(\tau,\beta)U(\tau,\beta)$,* *where $U(t,\beta)$ and $[\Psi(t,\beta)] = [\Psi_{jk}(t,\beta)]$ are independent (vector and matrix valued) mean zero Gaussian process with covariance given by:*

$$\begin{aligned} \operatorname{cov}[U(s,\beta),U(t,\beta)] &= \Sigma(s\wedge t,\beta), \\ \operatorname{cov}[\Psi_{jk}(s,\beta),\Psi_{m\ell}(t,\beta)] &= \gamma_{jk;m\ell}(s\wedge t,\beta), \end{aligned}$$

where Σ and γ are defined in Appendix B by B-4 and B-7, respectively.

If we define $\eta^{\star}_{jk}(t,\beta) = \eta_{jk}(t,\beta) + \hat{A}_{jk;0}(t,\hat{\beta})\mathbf{Z}_{jk}$, by Taylor expansion we can derive the asymptotics for $A_{jk}(t;\mathbf{Z}_{jk})$:

Lemma B.2 *$n^{1/2}(\hat{A}_{jk}(t,\hat{\beta};\mathbf{Z}_{jk}) - A_{jk}(t,\beta;\mathbf{Z}_{jk}))$ has the same asymptotic distribution of*

$$e^{\hat{\beta}^{\top}\mathbf{Z}_{jk}}\left(\Psi_{jk}(t,\beta) + (\eta^{\star}_{jk})^{\top}(t,\beta)\Sigma^{-1}(\tau,\beta)U(\tau,\beta)\right),$$

where $U(t,\beta)$ and $[\Psi(t,\beta)] = [\Psi_{jk}(t,\beta)]$ are the same processes of Proposition B.1.

Hence, in analogy of Gill (1980), for an individual with covariate $\mathbf{Z}$, the semi–Markov kernel $\mathbf{Q}(\tau;\mathbf{Z}_{jk})$ can be written in terms of product integrals:

$$Q_{jk}(\tau;\mathbf{Z}) = \int_0^{\tau}\left(\prod_{[0,s)}(1 - dA_j(s;\mathbf{Z}))\right)dA_{jk}(s;\mathbf{Z}_{jk}), \qquad \text{(B-3)}$$

where $A_j(s;\mathbf{Z}) = \sum_{\ell} A_{j\ell}(s;\mathbf{Z}_{j\ell})$. Under continuity conditions, (B-3) can be rewritten:

$$Q_{jk}(\tau;\mathbf{Z}) = \int_0^{\tau}\exp\{-A_j(u;\mathbf{Z})\}\,dA_{jk}(u;\mathbf{Z}_{jk}).$$

In the next section, the large sample theory of the empirical estimator suggested by (B-3) will be derived. This derivation is essentially based on the results contained in Dabrowska (1995).

B.2 Asymptotic distribution of semi–Markov kernel

In this section we recall the results from Dabrowska (1995) about the asymptotic covariance matrix f or the semi–Markov kernel, adapted to our context. We also recall the notation used and the following definitions, needed for stating the theorem in case of the semi–parametric proportional hazards model.

Given $S^{(0)}_{jk}(t,\beta)$, defined in (B-2), we denote with $S^{(1)}_{jk}(t,\beta)$ and $S^{(2)}_{jk}(t,\beta)$, the vector and respectively the matrix of the first and second partial derivatives of $S^{(0)}_{jk}(t,\beta)$ with respect to β. We define, in a neighborhood of the true parameter value β_0:

$$s^{(p)}_{jk}(t,\beta) := \mathbb{E}(S^{(p)}_{jk}(t,\beta)),$$

for $p = 0,1,2$. Moreover, set

$$e_{jk}(t,\beta) := \frac{s^{(1)}_{jk}(t,\beta)}{s^{(0)}_{jk}(t,\beta)},$$

$$v_{jk}(t,\beta) := \frac{s^{(2)}_{jk}(t,\beta)}{s^{(0)}_{jk}(t,\beta)} - e_{jk}(t,\beta)^{\otimes 2}$$

and define

$$\Sigma(t,\beta) := \sum_{j,k} \int_0^t v_{jk}(u,\beta)\, s^{(0)}_{jk}(u,\beta)\, \alpha_{jk;0}(u) du. \tag{B-4}$$

All the quantities so far recalled are commonly used in the proportional hazards setting. Furthermore, we want to introduce quantities specifically used in semi–Markov framework. Let us define:

$$\eta_{jk}(t) := -\int_{(0,t]} e_{jk}(u,\beta))dA_{jk;0}(u), \tag{B-5}$$

$$\eta^{\star}_{jk}(t,\mathbf{Z}_{jk}) := \eta_{jk}(t) + A_{jk;0}(t)\mathbf{Z}_{jk}, \tag{B-6}$$

$$\gamma_{jk;c\ell}(t,\beta) := \delta_{j,c}\,\delta_{k,\ell} \int_{(0,t]} \left(s^{(0)}_{jk}(u,\beta)\right)^{-1} \alpha_{0,jk}(u)\, du. \tag{B-7}$$

Proposition B.3 *Let τ_j be a point such that $\mathbb{E}Y_j(\tau) > 0$, $j = 1,\dots,r$ and suppose that $\mathbb{E}Y_j(0)^3 < \infty$ and the semi–Markov kernel $\mathbf{Q}$ continuous. Then the process $\sqrt{n}\,(\widehat{\boldsymbol{Q}_{jk}}(\cdot|\mathbf{Z}) - \boldsymbol{Q}_{jk}(\cdot|\mathbf{Z}))$ converges weakly in $D[0,\tau]^{r\times r}$ to mean zero Gaussian process $\mathbf{X}$, with covariance*

$$\mathbb{E}(X_{jk}(t)X_{lm}(s)) = e^{2\beta_0'\mathbf{Z}} \left(C_1(X_{jk}(t),X_{lm}(s)) + C_2(X_{jk}(t),X_{lm}(s))\right) \tag{B-8}$$

with

$$C_1(X_{jk}(s),X_{gh}(t)) := \sum_{\substack{q_1,p_1,\\ q_2}} \Delta_{p_1q_1}(k)\,\Delta_{p_1q_2}(k) \Big[Q_{jq_1}(t\wedge s) \int_0^{s\wedge t} dQ_{jq_2}(s_1)\,\gamma_{jp_1,jp_1}(s_1)$$

$$-\int_0^{s\wedge t} dQ_{jq_2}(s_1)\, Q_{jq_1}(s_1) \gamma_{ip_1,ip_1}(s_1) + Q_{jq_2}(t\wedge s) \int_0^{s\wedge t} dQ_{jq_1}(s_1)\,\gamma_{jp_1,jp_1}(s_1)$$

$$
\begin{aligned}
&-\int_0^{s\wedge t} dQ_{jq_1}(s_1)\,Q_{jq_2}(s_1)\gamma_{jp_1,jp_1}(s_1)\\
&+\mathbf{1}(t>s)(Q_{jp_1}(t)-Q_{jp_1}(s))\int_0^s dQ_{jq_2}(s_1)\gamma_{jp_1,jp_1}(s_1)\\
&+\mathbf{1}(t<s)(Q_{jq_2}(s)-Q_{jq_2}(t))\int_0^s dQ_{jp_1}(s_1)\gamma_{jp_1,jp_1}(s_1)\Bigg],
\end{aligned}
$$

and

$$
\begin{aligned}
C_2(X_{jk}(t),X_{lm}(s)) \;&:=\\
&\left[\sum_{p_1q_1}\Delta_{p_1q_1}(k)\int_0^t dQ_{jq_1}(s_1)\eta^{\star}_{jp_1}(s_1)\right]^T \Sigma^{-1}\left[\sum_{p_2q_2}\Delta_{p_2q_2}(m)\int_0^s dQ_{lq_2}(s_2)\eta^{\star}_{lp_2}(s_2)\right],
\end{aligned}
$$

where

$$\Delta_{ab}(c) := \delta_{ac}-\delta_{bc}.$$

For uniformity with the notation used in Appendix A for the non–parametric case, we denote with $E^{(g,h)}_{(j,k)}(s,t)$ the covariance matrix:

$$E^{(g,h)}_{(j,k)}(s,t) := \mathbb{E}(X_{jk}(s)X_{gh}(t)). \tag{B-9}$$

C Asymptotic variance for the transition probabilities

The variance calculation can be split in the terms:

$$\mathrm{var}(D_{jk}(t)) = \mathrm{var}(D^{(a)}_{jk}(t)) + \mathrm{var}(D^{(b)}_{jk}(t)) - 2\,\mathrm{cov}(D^{(a)}_{jk}(t)\,D^{(b)}_{jk}(t)). \tag{C-1}$$

For the first term, we can write:

$$
\begin{aligned}
\mathrm{var}(D^{(a)}_{jk}(t)) \quad = \quad & \sum_{m_1,m_2}\sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2}\mathbb{E}\Big(\big[\mathbf{Q}^{(m_1-k_1)}*\mathbf{X}*\mathbf{Q}^{(k_1-1)}*\mathbf{h}(t)\big]_{jk}\\
&\times\big[\mathbf{Q}^{(m_2-k_2)}*\mathbf{X}*\mathbf{Q}^{(k_2-1)}*\mathbf{h}(t)\big]_{jk}\Big)
\end{aligned}
$$

$$=: \quad \sum_{m_1,m_2} \sum_{k_1=1}^{m_1} \sum_{k_2=1}^{m_2} \mathbf{L}_{jk}(m_1,k_1;m_2,k_2). \tag{C-2}$$

Hence, integrating by parts and using the fact that for any $m > 0$, $\mathbf{Q}^{(m)}(0) = 0$ and $\mathbf{X}(0) = 0$, for the each term of the sum we have:

$$\begin{aligned}
& \mathbf{L}_{jk}(m_1,k_1;m_2,k_2) \\
= & \sum_{\substack{a_1,b_1,\\ a_2,b_2}} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \\
& \times \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \int_0^t du_2 \int_0^{t-u_2} Q_{j,a_2}^{(m_2-k_2)}(ds_2) \\
& \times \int_0^{t-u_2-s_2} Q_{b_2,k}^{(k_2-1)}(dt_2) \mathbf{h}_k(t-u_1)\, \mathbf{h}_k(t-u_2)\, \mathbb{E}\Big(X_{a_1,b_1}(t_1) X_{a_2,b_2}(t_2) \Big) \\
= & \sum_{\substack{a_1,b_1,\\ a_2,b_2}} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \\
& \times \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \int_0^t du_2 \int_0^{t-u_2} Q_{j,a_2}^{(m_2-k_2)}(ds_2) \\
& \times \int_0^{t-u_2-s_2} Q_{b_2,k}^{(k_2-1)}(dt_2) \mathbf{h}_k(t-u_1)\, \mathbf{h}_k(t-u_2)\, E_{(a_1,b_1)}^{(a_2,b_2)}(t_1,t_2)),
\end{aligned} \tag{C-3}$$

where in the last step we used (A-20) for evaluating $\mathbb{E}(X_{a_1,b_1}(t_1)X_{a_2,b_2}(t_2))$. Thus inserting (C-3) in (C-2) we have $\mathrm{var}(\mathbf{D}^{(a)}(t))$.

As regards $\mathrm{var}(D_{jk}^{(b)}(t))$, we can write:

$$\begin{aligned}
\mathrm{var}(D_{jk}^{(b)}(t)) &= \sum_{\ell_1,\ell_2} \int_0^t R_{jk}(ds_1) \int_0^t R_{jk}(ds_2) \mathbb{E}(X_{k\ell_1}(t-s_1)\, X_{k\ell_2}(t-s_2)) \\
&= \sum_{\ell_1,\ell_2} \int_0^t R_{jk}(ds_1) \int_0^t R_{jk}(ds_2) E_{(k,k)}^{(\ell_1,\ell_2)}(t-s_1,t-s_2),
\end{aligned} \tag{C-4}$$

where in the last step we apply the results of Appendix A or B in case of non–parametric or semi–parametric models respectly.

As regards the last term:

$$\begin{aligned}
\mathbb{E}(D_{jk}^{(a)}(t)\, D_{jk}^{(b)}(t)) \quad &= \quad \sum_{m_1,m_2} \sum_{k_1=1}^{m_1} \mathbb{E}\Big(\big[\mathbf{Q}^{(m_1-k_1)} * \mathbf{X} * \mathbf{Q}^{(k_1-1)} * \mathbf{h}(t)\big]_{jk} \\
& \qquad \times \big[\mathbf{Q}^{(m_2)} * \tilde{\mathbf{X}}\big]_{jk} \Big)
\end{aligned}$$

$$=: \quad \sum_{m_1,m_2} \sum_{k_1=1}^{m_1} \mathbf{M}_{jk}(m_1,k_1;m_2). \tag{C-5}$$

Similarly to the estimate (C-3), we have:

$$\begin{aligned}
\mathbf{M}_{jk}(m_1,k_1;m_2) &= \sum_{a_1,b_1} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \\
&\quad \times \int_0^t Q_{i,k}^{(m_2)}(ds_2) \mathbb{E}\Big(X_{a_1,b_1}(t_1) \tilde{X}_{kk}(t-s_2) \Big) \mathbf{h}_k(t-u_1) \\
&= \sum_{a_1,b_1,\ell} \int_0^t du_1 \int_0^{u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \int_0^{s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \\
&\quad \times \int_0^t Q_{j,k}^{(m_2)}(ds_2) \mathbb{E}\Big(X_{a_1,b_1}(t_1) X_{k\ell}(t-s_2) \Big) \mathbf{h}_k(t-u_1) \\
&= \sum_{a_1,b_1,\ell} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \\
&\quad \times \int_0^t Q_{j,k}^{(m_2)}(ds_2) \mathbb{E}\Big(X_{a_1,b_1}(t_1) X_{k\ell}(t-s_2) \Big) \mathbf{h}_k(t-u_1) \\
&= \sum_{a_1,b_1,\ell} \int_0^t du_1 \int_0^{t-u_1} Q_{j,a_1}^{(m_1-k_1)}(ds_1) \int_0^{t-u_1-s_1} Q_{b_1,k}^{(k_1-1)}(dt_1) \\
&\quad \times \int_0^t Q_{j,k}^{(m_2)}(ds_2) E_{(a_1,k)}^{(b_1,\ell)}(t_1,t-s_2)\, \mathbf{h}_k(t-u_1).
\end{aligned} \tag{C-6}$$

Therefore, inserting (C-2), (C-5), (3.19), in (C-1), we obtain the final result for the variance of the process.

References

Aalen, O. O., Johansen, S. (1978): An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.*, **5**(3), 141–150.

Andersen, P. K., (2002): Multi–state models for event history analysis. *Stat. Methods Med. Res.* **11**(2), 91-115.

Andersen, P. K., Borgan, Ø., Gill R. D., Keiding, N. (1993): *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 767.

Beyersmann, J., Allignol, A., Schumacher, M. (2012): *Competing risks and multi-state models with R*, Springer Verlag, 2012.

Çinlar, E. (1969): Markov renewal theory. *Advances in Appl. Probability*, **1**, 123–187.

Craddock, C., Szydlo, R. M., Klein, J. P., Dazzi, F., Olavarria, E., van Rhee, F., Pocock, C., Cwynarski, K., Apperley, J. F., Goldman, J. M.. (2000): Estimating leukemia-free survival after allografting for chronic myeloid leukemia: a new method that takes into account patients who relapse and are restored to complete remission. *Blood*, **96**(1), 86–90.

Dabrowska, D. M. (1995): Estimation of transition probabilities and bootstrap in a semiparametric Markov renewal model. *J. Nonparametr. Statist.*, **5**(3), 237–259.

Dabrowska, D. M., Sun, G. W., Horowitz, M. M. (1994): Cox regression in a markov renewal model: an application to the analysis of bone marrow transplant data. *J. Amer. Statist. Assoc.*, **89**(427), 867–877.

de Wreede, L. C., Fiocco, M., Putter, H. (2010): The mstate package for estimation and prediction in non– and semi–parametric multi–state and competing risks models. *Comput. Methods Programs Biomed.*, **99**(3), 261–274.

Gill, R. D. (1980): Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrsch. Verw. Gebiete*, **53**(1), 97–116.

Gill, R. D. (1989): Non– and semi–parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Statist.*, **16**(1),297–128.

Jackson, C. (2011): Multi-state models for panel data: the msm package for R. *J. Stat. Soft.*, **38**(8), 1–28.

Kang, M., Lagakos, S. W. (2007): Statistical methods for panel data from a semi-markov process, with application to HPV. *Biostatistics*, **8**(2), 252–264.

Lagakos, S. W., Sommer, C. J., Zelen, M. (1978): Semi-Markov models for partially censored data. *Biometrika*, **65**(2), 311–317.

Lin, D. Y., Fleming, T. R., Wei, L. J. (1994): Confidence bands for survival curves under the proportional hazards model. *Biometrika*, **81**(1), 73–81.

Mammen, E. (1992): Bootstrap, wild bootstrap, and asymptotic normality. *Probab. Theory Relat. Fields*, **93**(4), 439–455.

Monteiro, A. A., Smirnov, G. V., Lucas, A. (2006): Non–parametric estimation for non–homogeneous semi–Markov processes: an application to credit risk. Tinbergen Institute Discussion Papers 06-024/2, Tinbergen Institute.

Putter, H., Fiocco, M., Geskus, R. B (2007): Tutorial in biostatistics: competing risks and multi-state models. *Stat. Med.*, **26**(11), 2389–2430.

Shu, Y., Klein, J. P., Zhang M. J. (2007): Asymptotic theory for the Cox semi–Markov illness–death model. *Lifetime Data Anal.*, **13**(1), 91–117.

Termorshuizen, F., Korevaar, J. C., Dekker, F. W., van Manen, J. G., Boeschoten, E. W., and Krediet, R. T. (2003): Hemodialysis and peritoneal dialysis: comparison of adjusted mortality rates according to the duration of dialysis: analysis of The Netherlands Cooperative Study on the Adequacy of Dialysis 2. *J. Am. Soc.*

Nephrol., **14**(11), 2851–2860.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 443.

Xu, J., Kalbfleisch, J. D., Tai B. (2010): Statistical analysis of illness–death processes and semi–competing risks data. *Biometrics.*, **66**(3), 716–725.