Universidade de Vigo

Testing markovianity in the three-state progressive model via future-past association

Mar Rodríguez-Girondo and Jacobo de Uña Álvarez

Report 10/01

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://eioweb.uvigo.es/
E-mail: depc05@uvigo.es

Universidade de Vigo

Testing markovianity in the three-state progressive model via future-past association

Mar Rodríguez-Girondo and Jacobo de Uña Álvarez

Report 10/01

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita: 

Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

Testing markovianity in the three-state progressive model via future-past association

Mar Rodríguez-Girondo[1] and Jacobo de Uña-Álvarez[1,2]

[1]SiDOR Research Group, University of Vigo. E-mail: margirondo@uvigo.es
[2]Department of Statistics and OR, University of Vigo. E-mail: jacobo@uvigo.es

Abstract

The three-state progressive model is a special multi-state model with important applications in Survival Analysis. It provides a suitable representation of the individual's history when an intermediate event (with a possible influence on the survival prognosis) is experienced before the main event of interest. Estimation of transition probabilities in this and other multi-state models is usually performed through the Aalen-Johansen estimator. However, Aalen-Johansen may be biased when the underlying process is not Markov. In this paper we provide a new approach for testing markovianity in the three-state progressive model. The new method is based on measuring the future-past association along time. This results in a deep inspection of the process which often reveals a non-markovian behaviour with different trends in the association measure. A test of significance for zero future-past association at each time point is introduced, and a significance trace is proposed accordingly. The finite sample performance of the test is investigated through simulations. We illustrate the new method through real data analysis.

Keywords Markov condition; Multi-state models; Kendall's Tau; Recurrent Events.

1 Introduction

In longitudinal medical studies, patients are observed over time and they typically may experience several events of interest through the following-up. The statistical analysis in such studies is often performed using multi-state models. A multi-state model is a model for a continuous time stochastic process allowing individuals to move among a finite number of states. In biomedical applications, the states might be based on clinical symptoms (e.g. bleeding episodes), biological markers (e.g. CD4 T-lymphocyte cell counts), some scale of the disease (e.g. stages of cancer or VIH infection), or a non-fatal complication in the course of the illness (e.g. cancer recurrence). The simplest form of a multi-state model is the mortality model, with states 'alive' and 'dead' and one possible transition. By splitting the 'alive' state into two transient states which are visited in a successive way (with no chance of coming back) we obtain the

Figure 1: Three-state progressive model

three-state progressive model (see Figure 1). This model is convenient when there exists an intermediate event (e.g. a recurrence) which may influence the survival prognosis. More involved multi-state models such as the illness-death model or the bivariate model are sometimes needed. See Hougaard (2000), Andersen and Keiding (2002), or Meira-Machado et al. (2009) for an introduction to this area.

A typical situation in which a k-state progressive model is useful is when analyzing recurrent event data, which arise when each individual may go through a well-defined event several times along his history. Then, the inter-event times are referred to as the gap times, and they are of course determined by the times at which the recurrences take place (i.e. the recurrence times). See Cook and Lawless (2007) for an up-to-date revision of statistical methods for recurrent event data. In the three-state progressive model, the interest is focused on a given couple of (successive) gap times. In the real data analyzed in Sections 2 and 3, these will be the time up to first recurrence and the time from first to second recurrence for bladder cancer patients. These data come from a cancer bladder study (Byar (1980)) conducted by the Veterans Administration Cooperative Urological Research Group. In this study, patients had superficial bladder tumors that were remove transurethrally. Many patients had multiple recurrences of tumors during the study, and new tumors were removed at each visit. Here we analyzed the n =85 individuals in the placebo and thiotepa treatment groups; these data are listed in Wei et al. (1989). Only the first two recurrence times Z and T (or the corresponding gap times $T_1 = Z$ and $T_2 = T - Z$) are considered here, so a three-state progressive model applies. Among the 85 patients, 47 relapsed at least once (45% of censoring on T_1) and, among these, 29 had another recurrence (38% of extra censoring).

Usually, it is assumed that the stochastic process under investigation satisfies the Markov condition, and hence a Markovian multi-state model is fitted to the data at hand. A process is Markov if, conditionally on the present, its future evolution is independent of the past, including the states previously visited and the transition times among them. This condition is sometimes violated. For example, Andersen et al. (2000) provide evidence on non-Markovianity in a study on mortality in liver cirrhosis; in this case, the Markov condition fails because the mortality is markedly increased shortly after undergoing a bleeding episode, which is taken as an intermediate event. The bladder cancer data example is also non-Markovian, because the risk of suffering a second recurrence is relatively high just after having the first one (see Section 2 for further details on this). One of the main arguments for checking the Markov condition is that the usual estimators (e.g. Aalen-Johansen transition probabilities) may

be systematically biased in non-Markovian situations (Meira-Machado et al. (2006)). Another argument is that the Markov condition gives information about the way in which the process evolves, being of interest for its own right. Hence, it is interesting to know how to explore markovianity and to construct powerful testing methods for the Markov assumption. This is the goal of the present work.

The rest of the paper is organized as follows. After giving some notation, in Section 2 we introduce a measure of future-past association in the three-state progressive model. Basic properties of this measure such as consistency and distributional convergence are established. We also investigate the finite sample performance through simulations. The association trace along time is used in the analysis of the bladder cancer data. In Section 3 a formal test for no association is proposed. We introduce a bootstrap resampling plan to approximate the null distribution of the test statistic. A simulation study shows the power of test at different time points. Also, a significance trace is constructed for the bladder cancer data. A related (but different) test statistic based on the association levels between the censored versions of the transition times is introduced, and its advantages are explored. Main conclussions and a final discussion follow in Section 4, while the technical proofs are deferred to the Appendix.

2 Measuring future-past association

Having in mind the three-state progressive model discussed in the Introduction, let Z and T be the sojourn time in State 1 and the time to reach State 3, respectively. The Markov condition can be formulated as the conditional independence between T (future) and Z (past) given $A_t = \{Z \leq t < T\}$ (being in State 2 at the 'present time' t) for each $t > 0$. We use as measure of association the Kendall's Tau:

$$\tau_t = p_{c,t} - p_{d,t}, \qquad t > 0,$$

where $p_{c,t}$ and $p_{d,t}$ are the probability of concordance and discordance, respectively, for two pairs (Z, T) falling on A_t; more explicitly,

$$p_{c,t} = 2 \int \int F_t(x^-, y^-) F_t(dx, dy), \qquad p_{d,t} = 2 \int \int U_t(x, y) F_t(dx, dy),$$

where $F_t(x, y) = P(Z \leq x, T \leq y | A_t)$ stands for the joint distribution function (df) of (Z, T) conditionally on A_t, and where $U_t(x, y) = P(Z > x, T < y | A_t) = F_t(t, y^-) - F_t(x, y^-)$. In the uncensored setup, this measure can be estimated by plugging-in the empirical df subject to $\{Z \leq t < T\}$. In most practical cases, however, the sample will be subject to right-censoring because follow-up time limitations, withdrawals, and so on. Let C be the censoring time, which we assume to be independent of (Z, T). Furthermore, let $\widetilde{Z} = \min(Z, C)$ and $\widetilde{T} = \min(T, C)$ be the censored versions of Z and T respectively, and let $\Delta_1 = I(Z \leq C)$ and $\Delta = I(T \leq C)$ be their corresponding censoring indicators. The

sample information is represented by $\left\{\left(\widetilde{Z}_i, \widetilde{T}_i, \Delta_{1i}, \Delta_i\right), i = 1, ..., n\right\}$, iid copies of $\left(\widetilde{Z}, \widetilde{T}, \Delta_1, \Delta\right)$. In practice, only situations $(\Delta_1, \Delta) = (0, 0)$, $(1, 0)$, and $(1, 1)$ may happen, corresponding to individuals censored in State 1, in State 2, or uncensored, respectively. Besides, we assume that the individuals observed to pass through State 2 coincide to those with $\widetilde{Z} < \widetilde{T}$ (in words: we exclude the possibility of a zero transition time from State 2 to State 3). On the basis of the censored information, we propose to estimate F_t through

$$\widehat{F}_t(x, y) = \frac{1}{\widehat{P}(A_t)} \sum_{i=1}^{n} W_{ni} I\left(\widetilde{Z}_{i:n} \leq x, \widetilde{T}_{i:n} \leq y\right) I(\widetilde{Z}_{i:n} \leq t < \widetilde{T}_{i:n})$$

where

$$W_{ni} = \frac{\Delta_{i:n}}{n - i + 1} \prod_{j=1}^{i-1} \left[1 - \frac{\Delta_{j:n}}{n - j + 1}\right], \qquad i = 1, ..., n,$$

$\widetilde{T}_{1:n} \leq ... \leq \widetilde{T}_{n:n}$ is the ordered $\widetilde{T}$-sample, and $\left(\widetilde{Z}_{i:n}, \Delta_{i:n}\right)$ is the i-the concomitant. Hence, W_{ni} is just the jump of the Kaplan-Meier estimator of the df of T at time $\widetilde{T}_{i:n}$, which satisfies $W_{ni} = 1/n$, $1 \leq i \leq n$, when there is no censoring. For the estimator $\widehat{P}(A_t)$ we propose to use $\widehat{P}(A_t) = \widehat{S}_T(t) - \widehat{S}_Z(t)$, where $\widehat{S}_T$ and $\widehat{S}_Z$ are the (standard) Kaplan-Meier estimators of the survival functions of T and Z respectively. Note that, in the uncensored case, $\widehat{P}(A_t)$ is just the sample proportion of individuals in State 2 at time t. The censored version of the Kendall's Tau is given consequently by $\widehat{\tau}_t = \widehat{p}_{c,t} - \widehat{p}_{d,t}$ where

$$\widehat{p}_{c,t} = 2 \int \int \widehat{F}_t(x^-, y^-) \widehat{F}_t(dx, dy), \qquad \widehat{p}_{d,t} = 2 \int \int \widehat{U}_t(x, y) \widehat{F}_t(dx, dy),$$

where $\widehat{U}_t(x, y) = \widehat{F}_t(t, y^-) - \widehat{F}_t(x, y^-)$.

In the following result we establish the consistency of $\widehat{\tau}_t$. This is of course a consequence of the consistency of $\widehat{F}_t$. The idea of weighting a multivariate distribution by marginal Kaplan-Meier weights goes back to Stute (1993). The proof in the Appendix shows that his result still applies here, by taking the Z time as a (one-dimensional) 'covariate'. Before stating the result, we need to introduce some further notation. Let $b_H = \inf\{t : H(t) = 1\}$, where H is the df of $\widetilde{T}$. Introduce

$$\tau_t^0 = p_{c,t}^0 - p_{d,t}^0$$

where

$$p_{c,t}^0 = 2 \int \int_{\{y \leq b_H\}} F_t(x^-, y^-) F_t(dx, dy), \qquad p_{d,t}^0 = 2 \int \int_{\{y \leq b_H\}} U_t(x, y) F_t(dx, dy),$$

which is just the Kendall's Tau subject to the restriction $\{T \leq b_H\}$. When the support of the censoring distribution contains that of T this restriction is

superfluous, and we have $\tau_t^0 = \tau_t$. However, in general the equality does not hold. Actually, one can not expect consistent estimation of τ_t when the censoring skips a relevant subset of the lifetime values. However, interestingly, when T and Z are independent conditionally on A_t, it happens that $\tau_t^0 = \tau_t (= 0)$, a fact that can be easily checked by using the relation $F_t(x, y) = F_t(x, \infty)F_t(\infty, y)$. Hence, departures from zero of $\hat{\tau}_t$ can be interpreted as a lack of markovianity of the process even when $P(T \leq b_H) < 1$.

Let F_T and G denote the df of T and C respectively.

Theorem 1 *Take $t \leq b_H$. If H is continuous, we have $\widehat{\tau}_t \to \tau_t^0$ with probability one. Furthermore, under $\int_0^{b_H} (1-G)^{-1} dF_T < \infty$, if $F_T(b_H) < 1$ or if $F_T(b_H) = 1$ and $G(b_H) < 1$, we have that $n^{1/2}\left(\widehat{\tau}_t - \tau_t^0\right)$ converges in law to a zero-mean normal random variable.*

Proof. See the Appendix.

Remark. In the case $F_T(b_H) = G(b_H) = 1$ condition $\int_0^{b_H} (1-G)^{-1} dF_T < \infty$ may not be enough to ensure convergence in distribution. Stute (1995) nicely reported the example in which $1-F_T \sim c(1-G)^\beta$ around b_H for some $c > 0, \beta > 0$; in such a case, when $\beta > 1$ one must rather impose $\int_0^{b_H} (1-G)^{-\alpha} dF_T < \infty$ where $\alpha = (1+\beta)/2$.

We have performed a small simulation study to investigate the accuracy of $\widehat{\tau}_t$ as an estimator of τ_t. Specifically, we have simulated 1,000 Monte Carlo trials of three different models. These models were based in the accelerated failure time specification $\log(T - Z) = f(Z) + \varepsilon$ where the error ε is independent of the 'covariate' Z, while $f(.)$ is the predictor. An alternative representation is given by $\lambda(t|s) = \lambda_0\left((t-s)\, e^{-f(s)}\right) e^{-f(s)}$ where $\lambda(t|s)$ is the conditional hazard of T given $Z = s$ and λ_0 stands for the hazard of $W = e^{\varepsilon}$. This formulation belongs to the proportional hazards family when ε follows a extreme-value distribution (i.e. λ_0 is constant). On the other hand, the Markov condition holds if and only if $\lambda(t|s)$ is free of s. The Z was distributed as a $U[0, 2]$ random variable and the models were as follows:

Model 0 (Markovian). λ_0 constant and $f(.) \equiv 0$

Model 1 (Non-Markovian, proportional hazards). λ_0 constant and $f(s) = (s-1)^2$

Model 2 (Non-Markovian, non-proportional hazards). λ_0 loglogistic and $f(.) \equiv 0$

The traces of Kendall's tau τ_t for Models 1 and 2 along the interval $t \in [0.5, 3]$ are displayed in Figure 2 (black lines). From this Figure we can see that Model 1 presents first a negative and then a positive future-past association, and that τ_t vanishes for $t \approx 1.8$. This is a consequence of the increasing-decreasing shape of the conditional hazard $\lambda(t|s) = \exp\left[-(s-1)^2\right]$ on the interval $[0, 2]$ (note that larger hazard values correspond to smaller T's). Model 2 presents a negative future-past association which is stronger as t grows; in this case, $\lambda(t|s)$ equals

the loglogistic hazard $\lambda_0(t-s) = (1+t-s)^{-1}$ which is monotone increasing in s, and hence the negative values of τ_t were expected. Of course, we omit the trace corresponding to Model 0 because in this case $\tau_t = 0$ for each t-value. The traces of τ_t^0 for Models 1 and 2 and the several censoring degrees are also included in Figure 2. It can be appreciated that τ_t^0 gets away from τ_t as the censoring increases, and that $\tau_t^0 = 0$ whenever $\tau_t = 0$. This is in agreement with our previous discussion on the limit of $\widehat{\tau}_t$.

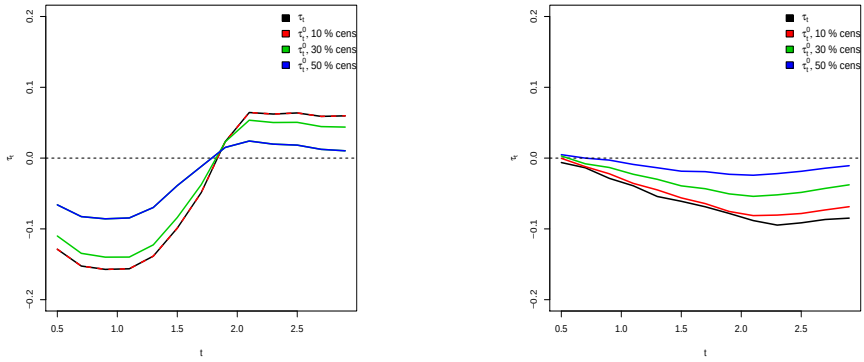

Figure 2: Comparision between τ_t and τ_t^0 for several censoring rates in non-markovian simulated models. Monte Carlo approximation based on 100,000 observations. Left panel: Model 1. Right panel: Model 2.

To evaluate the performance of $\widehat{\tau}_t$, a grid of t-values and different sample sizes and censoring percentages were considered. Specifically, we took $t = 0.7$, 1.1, 1.5, 1.9, 2.3, 2.7, $n = 100$, 250, and 500, and censoring levels of 0, 10, 30 and 50%. For the censoring distribution G, we took a Uniform with support $[0, b_G]$, where b_G was chosen to obtain the desired censoring level. Biases and standard deviations of $\widehat{\tau}_t$ along the 1,000 trials are given in Tables 1 (Model 0), 2 (Model 1), and 3 (Model2). From these Tables we see that the standard deviations decrease as the sample size increases, and that (generally speaking) larger dispersions were obtained with heavy censoring. We also see that the standard deviations are larger for small and (specially) large t's; this is because $\widehat{\tau}_t$ is estimating the future-past association for the individuals in State 2 at time t, and the information is naturally scarcer at these points. Indeed, for Model 0, $n = 100$, and $t \geq 2.3$, we could not compute the estimator because we had no data. The bias was of a smaller order with respect to the standard deviation, except for Models 1 and 2 with 50% of censoring; in these cases, there exists a systematic bias which does not vanishes as $n \to \infty$, because the limit τ_t^0 and the target τ_t are clearly different (note also the changing in the sign of the bias along the grid of t-values for Model 1, according to the black line in Figure 2, left). In summary, the behaviour of $\widehat{\tau}_t$ as an estimator of τ_t is reasonably good.

Table 1: Model 0 (Markovian). Bias and standard deviation of $\widehat{\tau}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n).

Censoring rate(%)	n	0.7 ($\tau_t = 0$)		1.1($\tau_t = 0$)		1.5($\tau_t = 0$)		1.9($\tau_t = 0$)		2.3($\tau_t = 0$)		2.7($\tau_t = 0$)	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
	100	-0.002	0.143	0.001	0.119	-0.004	0.111	-0.003	0.108	-0.000	0.122	-0.002	0.157
0	250	-0.002	0.085	0.000	0.077	-0.003	0.067	0.003	0.066	0.004	0.074	0.003	0.093
	500	0.003	0.059	0.000	0.053	0.002	0.048	-0.000	0.047	-0.000	0.054	-0.004	0.064
	100	-0.002	0.128	0.001	0.113	-0.003	0.109	-0.002	0.117	-0.000	0.137	0.004	0.182
10	250	-0.001	0.076	0.001	0.070	-0.004	0.067	0.003	0.073	0.002	0.081	0.002	0.104
	500	0.001	0.053	-0.000	0.048	0.002	0.047	-0.000	0.050	-0.000	0.058	-0.004	0.071
	100	-0.000	0.113	0.001	0.106	-0.004	0.112	-0.005	0.146	-0.009	0.190	-0.008	0.346
30	250	-0.001	0.068	-0.000	0.064	-0.004	0.074	0.003	0.095	-0.000	0.116	0.003	0.154
	500	-0.000	0.048	-0.000	0.048	-0.000	0.053	-0.000	0.065	-0.001	0.082	-0.003	0.106
	100	0.000	0.111	-0.001	0.107	-0.006	0.115	-0.008	0.163				
50	250	-0.000	0.069	-0.005	0.067	-0.002	0.085	0.002	0.115	0.002	0.150	0.005	0.178
	500	-0.000	0.047	-0.000	0.052	0.002	0.061	0.002	0.083	-0.000	0.110	-0.005	0.130

Table 2: Model 1 (Non-Markovian). Bias and standard deviation of $\widehat{\tau}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n).

Censoring rate(%)	n	0.7($\tau_t = -0.152$)		1.1($\tau_t = -0.156$)		1.5($\tau_t = -0.099$)		1.9($\tau_t = 0.023$)		2.3($\tau_t = 0.062$)		2.7($\tau_t = 0.059$)	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
	100	0.000	0.127	0.000	0.113	-0.006	0.105	0.001	0.104	-0.002	0.115	0.006	0.134
0	250	0.004	0.081	-0.001	0.071	-0.008	0.064	0.001	0.062	0.004	0.070	0.010	0.078
	500	0.006	0.055	0.001	0.048	-0.002	0.046	0.001	0.045	0.004	0.049	0.008	0.055
	100	0.020	0.120	0.016	0.112	0.004	0.108	0.000	0.115	-0.006	0.131	0.001	0.153
10	250	0.024	0.079	0.017	0.071	0.001	0.066	0.001	0.071	0.003	0.079	0.008	0.091
	500	0.024	0.054	0.016	0.048	0.005	0.047	0.001	0.050	0.003	0.057	0.006	0.065
	100	0.068	0.123	0.059	0.114	0.032	0.122	-0.004	0.148	-0.024	0.176	-0.018	0.203
30	250	0.063	0.079	0.055	0.076	0.024	0.080	-0.006	0.099	-0.018	0.116	-0.016	0.133
	500	0.064	0.056	0.053	0.052	0.031	0.059	0.000	0.071	-0.012	0.083	0.012	0.097
	100	0.114	0.117	0.107	0.116	0.067	0.133	-0.022	0.153	-0.056	0.174	-0.048	0.193
50	250	0.111	0.079	0.104	0.075	0.066	0.087	-0.013	0.107	-0.044	0.121	-0.045	0.131
	500	0.109	0.055	0.107	0.055	0.069	0.064	-0.009	0.082	-0.041	0.094	-0.045	0.098

In Figure 4 (top left panel) we depict the estimated future-past association trace for the bladder cancer data. Because of the presence of ties, the discrete version of the Kendall's Tau

$$\overline{\tau}_t = \frac{\widehat{\tau}_t}{\widehat{p}_{c,t} + \widehat{p}_{d,t}} = \frac{\widehat{p}_{c,t} - \widehat{p}_{d,t}}{\widehat{p}_{c,t} + \widehat{p}_{d,t}}$$

was used. Confidence limits (at 95%) based on the simple bootstrap are also displayed. This Figure 4 (top left panel) suggests a negative future-past association around $t = 3$ and $t = 9$, meaning that the risk of having a second recurrence is higher just after suffering the first one (more evidence on this will be reported in the next Section). Interestingly, this agrees with the analysis of markovianity based on the proportional hazards model $\lambda(t|s) = \lambda_0(t)e^{\beta s}$ under which $\widehat{\beta} = 0.0956$ ($s.e. = 0.0351$) with a logrank test's p-value of 0.00372. The proportional hazards approach has been traditionally used to explore markovianity in multi-state models (e.g. Kay (1986)) because of its simplicity. However, it is not an omnibus method; see our simulations in Section 3 for illustration of this fact.

Table 3: Model 2 (Non-Markovian). Bias and standard deviation of $\widehat{\tau}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n).

Censoring rate(%)	n	$0.7(\tau_t = -0.014)$		$1.1(\tau_t = -0.039)$		$1.5(\tau_t = -0.061)$		$1.9(\tau_t = -0.078)$		$2.3(\tau_t = -0.095)$		$2.7(\tau_t = -0.087)$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
0	100	0.000	0.134	-0.004	0.113	-0.002	0.103	0.002	0.091	0.007	0.103	0.005	0.118
	250	0.000	0.082	0.002	0.067	0.002	0.063	-0.001	0.059	0.001	0.065	-0.002	0.074
	500	-0.004	0.056	0.001	0.047	0.002	0.042	0.000	0.040	0.004	0.044	0.001	0.050
10	100	0.005	0.133	0.004	0.114	0.008	0.105	0.013	0.100	0.019	0.112	0.015	0.127
	250	0.007	0.083	0.010	0.071	0.011	0.066	0.012	0.065	0.014	0.073	0.014	0.084
	500	-0.001	0.057	0.007	0.049	0.009	0.046	0.009	0.047	0.014	0.053	0.011	0.060
30	100	0.012	0.124	0.020	0.112	0.028	0.107	0.040	0.113	0.050	0.132	0.046	0.153
	250	0.011	0.077	0.024	0.072	0.031	0.071	0.039	0.080	0.045	0.091	0.046	0.102
	500	0.006	0.058	0.020	0.055	0.027	0.054	0.032	0.057	0.041	0.066	0.039	0.073
50	100	0.014	0.114	0.032	0.109	0.049	0.107	0.063	0.127	0.075	0.143	0.072	0.153
	250	0.014	0.068	0.034	0.066	0.048	0.066	0.063	0.080	0.073	0.089	0.072	0.095
	500	0.013	0.051	0.032	0.048	0.049	0.053	0.062	0.061	0.075	0.071	0.072	0.074

Theorem 1 allows in principle for the construction of a α-level rejection region for the null hypothesis $H_{0t} : \tau_t = 0$. Note that the Markov condition is represented by the intersection null $H_0 = \bigcap_{t>0} H_{0t}$, so evidence against any specific H_{0t} can be interpreted as a lack of markovianity of the process. In practice, however, the variance of the asymptotic normal law must be estimated, which can be a difficult issue. For this reason, in Section 3 we introduce the bootstrap as a method to approximate the null distribution of $\widehat{\tau}_t$. For this aim, the simple bootstrap considered in Figure 4 (top left panel) will be no longer valid, since the information of the null hypothesis must be incorporated to the resampling plan.

3 Testing for no association

In this Section we formally introduce a method to test for the null hypothesis $H_{0t} : \tau_t = 0$ for each fixed time point t. The basic idea is to reject H_{0t} for large values of $|\widehat{\tau}_t|$. Given the significance level α for the test, the critical region $\{|\widehat{\tau}_t| > c_{t,\alpha}\}$, where $c_{t,\alpha}$ satisfies $P\left(|\widehat{\tau}_t| > c_{t,\alpha} \ / \ H_{0t}\right) = \alpha$, is approximated by bootstrapping. In order to include the null hypothesis in the resampling plan, we propose to draw the Z's independently of the T's for each fixed subpopulation $\{Z \leq t < T\}$. More specifically, the bootstrap resampling plan is as follows:

Step 1. Draw Z_i^* from the empirical df of the $\widetilde{Z}_i$'s such that $\widetilde{Z}_i \leq t < \widetilde{T}_i$; that is, $P^*(Z_i^* = \widetilde{Z}_{i:n}) = I(\widetilde{Z}_{i:n} \leq t < \widetilde{T}_{i:n})/n_t$ where $n_t = \sum_{j=1}^n I(\widetilde{Z}_{j:n} \leq t < \widetilde{T}_{j:n})$.

Step 2. Draw independently T_i^* from $\widehat{F}_t(\infty, .)$; that is, $P^*(T_i^* = \widetilde{T}_{i:n}) = W_{ni} I(\widetilde{Z}_{i:n} \leq t < \widetilde{T}_{i:n})/\widehat{P}(A_t)$

Step 3. Draw independently C_i^* from $\widehat{G}_t(.)$, the Kaplan-Meier estimator of $G_t(y) = P(C \leq y / C > t)$ based on the $\left(\widetilde{T}_i, 1 - \Delta_i\right)$'s.

Step 4. Set $\widetilde{Z}_i^* = Z_i^*$, $\Delta_{1i}^* = 1$, $\widetilde{T}_i^* = \min(T_i^*, C_i^*)$, $\Delta_i^* = I\left(T_i^* \leq C_i^*\right)$.

Note that in Step 1 we are using ordinary (rather than Kaplan-Meier) weights in the resampling. This is because the cases with $\widetilde{Z}_i \leq t < \widetilde{T}_i$ report uncensored values for $\widetilde{Z}_i$. This also explains why we are setting $\Delta^*_{1i} = 1$ in Step 4. We repeat Steps 1-4 until a bootstrap resample of size n_t, say $\left\{\left(\widetilde{Z}^*_i, \widetilde{T}^*_i, \Delta^*_{1i}, \Delta^*_i\right), i = 1, ..., n_t\right\}$, is obtained. The whole procedure is repeated a large number B of times to get B bootstrap resamples; the corresponding values of the censored Kendall's Tau, $\widehat{\tau}^{*b}_t, b = 1, ..., B$, are then obtained. We then define $c^*_{t,\alpha}$ as the $100(1-\alpha)\%$ percentile of the $\left|\widehat{\tau}^{*b}_t\right|$'s.

In Tables 4 to 6 we report the rejection proportion of the introduced test along a grid of t-values, for $M = 1,000$ Monte Carlo trials from Models 0, 1 and 2 above. A significance level of 5% was used, and we took $B = 200$. We took a relatively low value for B because of the high computational cost involved (note that for g points in the t-grid we are taking MBg bootstrap resamples). In any case, we have tried larger values of B and the obtained results are similar to those provided in Tables 4 to 6 (results not shown). The same sample sizes and censoring degrees as in Tables 1 to 3 were considered here. Note that Model 0 is Markovian, so we expect rejection proportions about 0.05 in this case. Results in Table 4 are quite satisfactory to this regard; we can appreciate however that the test becomes conservative when the censoring grows (30, 50%), particularly at the left of the t-grid. For Models 1-2 we expect a rejection proportion increasing with the sample size and also with the proportion of uncensoring. These features are appreciated in Tables 5 and 6. Interestingly, we see that the maximum power is achieved at some central point in the grid ($t = 1.1$ in Model 1, $t = 2.3$ in Model 2). This is a consequence of two facts. First, the amount of information on τ_t grows at central points; second, we have that the largest absolute value of τ_t (see the black lines in Figure 2), and hence the alternative most separated from the null, is obtained precisely for $t = 1.1$ (Model 1) or $t = 2.3$ (Model 2). It is also remarkable the lost of power of the Kendall's tau in Model 1 for $t = 1.9$; this is because the level of association is zero at this point. So in practice the value of t may have a big impact in the power of the test.

For comparison purposes, the rejection proportion corresponding to the simple method based on a proportional hazard specification with linear predictor, $\lambda(t|s) = \lambda_0(t)e^{\beta s}$, is included in Tables 4 to 6. It is interesting to see how the power of this test may not increase with the sample size, as it happens in Table 5 (Model 1). In this case, the reason is found in the miss-specification of the predictor, which can not detect the parabolic influence $f(s) = (s-1)^2$. A bit surprisingly, more power is reached with larger censoring proportions; this is due to the shortening in the observable support of the Z variable, which transforms the decreasing-increasing shape of the predictor into a fairly monotone decreasing curve. The situation changes in Table 6, where (according to the intuition) the power of this testing strategy decreases as the censoring grows. In any case, the method based on the Kendall's tau gave more power than the proportional hazards based method for particular values of t (Tables 5 and 6).

Table 4: Model 0 (Markovian). Rejection proportions of tests based on $\hat{\tau}_t$ and $\hat{\hat{\tau}}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

Censoring rate(%)	n	Statistic	0.7	1.1	1.5	1.9	2.3	2.7	PH method
	100	$\hat{\tau}_t$	0.046	0.048	0.048	0.058	0.042	0.063	0.053
		$\hat{\hat{\tau}}_t$	0.046	0.048	0.058	0.042	0.063	0.053	
0	250	$\hat{\tau}_t$	0.043	0.051	0.047	0.051	0.044	0.048	0.048
		$\hat{\hat{\tau}}_t$	0.043	0.051	0.047	0.051	0.044	0.048	
	500	$\hat{\tau}_t$	0.043	0.053	0.053	0.063	0.046	0.043	0.057
		$\hat{\hat{\tau}}_t$	0.043	0.053	0.053	0.063	0.046	0.043	
	100	$\hat{\tau}_t$	0.046	0.048	0.048	0.058	0.042	0.063	0.053
		$\hat{\hat{\tau}}_t$	0.049	0.048	0.052	0.046	0.058	0.058	
10	250	$\hat{\tau}_t$	0.036	0.044	0.042	0.048	0.040	0.041	0.045
		$\hat{\hat{\tau}}_t$	0.047	0.061	0.041	0.050	0.044	0.039	
	500	$\hat{\tau}_t$	0.038	0.030	0.046	0.059	0.046	0.037	0.056
		$\hat{\hat{\tau}}_t$	0.047	0.044	0.054	0.066	0.048	0.039	
	100	$\hat{\tau}_t$	0.017	0.016	0.022	0.049	0.06	0.059	0.053
		$\hat{\hat{\tau}}_t$	0.049	0.044	0.044	0.057	0.059	0.059	
30	250	$\hat{\tau}_t$	0.019	0.019	0.025	0.048	0.047	0.052	0.046
		$\hat{\hat{\tau}}_t$	0.043	0.048	0.045	0.053	0.050	0.049	
	500	$\hat{\tau}_t$	0.028	0.023	0.035	0.044	0.044	0.045	0.058
		$\hat{\hat{\tau}}_t$	0.050	0.051	0.048	0.053	0.041	0.040	
	100	$\hat{\tau}_t$	0.015	0.011	0.010	0.049			0.050
		$\hat{\hat{\tau}}_t$	0.055	0.063	0.044	0.052			
50	250	$\hat{\tau}_t$	0.011	0.008	0.015	0.042	0.043	0.043	0.048
		$\hat{\hat{\tau}}_t$	0.041	0.040	0.045	0.057	0.039	0.048	
	500	$\hat{\tau}_t$	0.010	0.022	0.017	0.036	0.047	0.030	0.047
		$\hat{\hat{\tau}}_t$	0.058	0.050	0.060	0.052	0.049	0.046	

Table 5: Model 1 (Non-Markovian). Rejection proportions of tests based on $\hat{\tau}_t$ and $\hat{\hat{\tau}}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

Censoring rate(%)	n	Statistic	0.7	1.1	1.5	1.9	2.3	2.7	Cox method
	100	$\hat{\tau}_t$	0.219	0.296	0.181	0.066	0.098	0.086	0.065
		$\hat{\hat{\tau}}_t$	0.219	0.296	0.181	0.066	0.098	0.086	
0	250	$\hat{\tau}_t$	0.449	0.622	0.379	0.072	0.171	0.146	0.071
		$\hat{\hat{\tau}}_t$	0.449	0.622	0.379	0.072	0.171	0.146	
	500	$\hat{\tau}_t$	0.738	0.890	0.620	0.100	0.288	0.247	0.087
		$\hat{\hat{\tau}}_t$	0.738	0.890	0.620	0.100	0.288	0.247	
	100	$\hat{\tau}_t$	0.169	0.243	0.150	0.075	0.111	0.091	0.073
		$\hat{\hat{\tau}}_t$	0.189	0.260	0.156	0.066	0.103	0.089	
10	250	$\hat{\tau}_t$	0.375	0.529	0.310	0.077	0.159	0.121	0.060
		$\hat{\hat{\tau}}_t$	0.394	0.540	0.323	0.073	0.155	0.133	
	500	$\hat{\tau}_t$	0.642	0.837	0.533	0.100	0.245	0.194	0.098
		$\hat{\hat{\tau}}_t$	0.667	0.835	0.542	0.091	0.248	0.190	
	100	$\hat{\tau}_t$	0.059	0.103	0.078	0.074	0.085	0.081	0.074
		$\hat{\hat{\tau}}_t$	0.135	0.188	0.111	0.069	0.069	0.075	
30	250	$\hat{\tau}_t$	0.143	0.239	0.160	0.086	0.101	0.082	0.073
		$\hat{\hat{\tau}}_t$	0.281	0.400	0.226	0.073	0.103	0.089	
	500	$\hat{\tau}_t$	0.297	0.469	0.246	0.113	0.127	0.102	0.118
		$\hat{\hat{\tau}}_t$	0.508	0.685	0.388	0.084	0.162	0.115	
	100	$\hat{\tau}_t$	0.017	0.027	0.041	0.049	0.037	0.059	0.061
		$\hat{\hat{\tau}}_t$	0.112	0.122	0.093	0.072	0.071	0.091	
50	250	$\hat{\tau}_t$	0.030	0.050	0.049	0.048	0.038	0.036	0.104
		$\hat{\hat{\tau}}_t$	0.192	0.265	0.143	0.063	0.067	0.065	
	500	$\hat{\tau}_t$	0.071	0.104	0.049	0.033	0.047	0.036	0.172
		$\hat{\hat{\tau}}_t$	0.371	0.471	0.228	0.070	0.103	0.071	

Table 6: Model 2 (Non-Markovian). Rejection proportions of tests based on $\hat{\tau}_t$ and $\hat{\tilde{\tau}}_t$ along 1,000 trials for several values of t, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

Censoring rate(%)	n	Statistic	0.7	1.1	1.5	1.9	2.3	2.7	Cox method
	100	$\hat{\tau}_t$	0.051	0.064	0.094	0.129	0.143	0.099	0.117
		$\hat{\tilde{\tau}}_t$	0.051	0.064	0.094	0.129	0.143	0.099	
0	250	$\hat{\tau}_t$	0.058	0.071	0.169	0.280	0.297	0.234	0.240
		$\hat{\tilde{\tau}}_t$	0.058	0.071	0.169	0.280	0.297	0.234	
	500	$\hat{\tau}_t$	0.053	0.113	0.248	0.478	0.531	0.400	0.410
		$\hat{\tilde{\tau}}_t$	0.053	0.113	0.248	0.478	0.531	0.400	
	100	$\hat{\tau}_t$	0.045	0.050	0.075	0.089	0.101	0.070	0.103
		$\hat{\tilde{\tau}}_t$	0.051	0.063	0.090	0.110	0.116	0.080	
10	250	$\hat{\tau}_t$	0.050	0.069	0.119	0.167	0.183	0.135	0.210
		$\hat{\tilde{\tau}}_t$	0.064	0.075	0.139	0.224	0.255	0.181	
	500	$\hat{\tau}_t$	0.038	0.095	0.183	0.322	0.342	0.246	0.373
		$\hat{\tilde{\tau}}_t$	0.047	0.115	0.237	0.415	0.466	0.348	
	100	$\hat{\tau}_t$	0.027	0.038	0.032	0.051	0.054	0.056	0.072
		$\hat{\tilde{\tau}}_t$	0.058	0.046	0.082	0.086	0.074	0.072	
30	250	$\hat{\tau}_t$	0.034	0.036	0.046	0.086	0.087	0.059	0.144
		$\hat{\tilde{\tau}}_t$	0.061	0.073	0.103	0.141	0.167	0.119	
	500	$\hat{\tau}_t$	0.036	0.040	0.079	0.115	0.135	0.091	0.265
		$\hat{\tilde{\tau}}_t$	0.050	0.093	0.183	0.287	0.286	0.196	
	100	$\hat{\tau}_t$	0.017	0.019	0.021	0.027	0.032	0.048	0.048
		$\hat{\tilde{\tau}}_t$	0.060	0.056	0.065	0.058	0.063	0.071	
50	250	$\hat{\tau}_t$	0.021	0.008	0.028	0.032	0.033	0.026	0.078
		$\hat{\tilde{\tau}}_t$	0.051	0.063	0.077	0.082	0.085	0.068	
	500	$\hat{\tau}_t$	0.018	0.022	0.034	0.054	0.054	0.041	0.120
		$\hat{\tilde{\tau}}_t$	0.045	0.071	0.107	0.139	0.128	0.088	

In the simulations above we have also included a test based in a different statistic, $\widehat{\widetilde{\tau}}_t$, which is defined as the (estimated) Kendall's tau between the censored versions of T ($\widetilde{T}$) and Z ($\widetilde{Z}$) given $\widetilde{A}_t = \left\{\widetilde{Z} \leq t < \widetilde{T}\right\}$ for each $t > 0$, that is, $\widehat{\widetilde{\tau}}_t$ is an estimator for $\widetilde{\tau}_t = \widetilde{p}_{c,t} - \widetilde{p}_{d,t}$ where

$$\widetilde{p}_{c,t} = 2\int\int \widetilde{F}_t(x^-, y^-)\widetilde{F}_t(dx, dy), \qquad \widetilde{p}_{d,t} = 2\int\int \widetilde{U}_t(x, y)\widetilde{F}_t(dx, dy),$$

where $\widetilde{F}_t(x, y) = P(\widetilde{Z} \leq x, \widetilde{T} \leq y|\widetilde{A}_t)$ stands for the joint df of $\left(\widetilde{Z}, \widetilde{T}\right)$ conditionally on $\widetilde{A}_t$, and where $\widetilde{U}_t(x, y) = \widetilde{F}_t(t, y^-) - \widetilde{F}_t(x, y^-)$. Note that $\widetilde{A}_t = A_t \cap \{C > t\}$ refers to the subpopulation being in State 2 at time t, which has not been censored by that time. As long as C is independent of the process, this subpopulation is representative of the subpopulation A_t. Put $\widetilde{H}_{0t} : \widetilde{\tau}_t = 0$. When T and Z are independent conditionally on A_t we have that $\widetilde{H}_{0t}$ holds. The converse is also true provided that the support of C contains that of T. This is stated as a Theorem.

Theorem 2 *If T and Z are independent conditionally on A_t, then $\widetilde{T}$ and $\widetilde{Z}$ are independent conditionally on $\widetilde{A}_t$. Conversely, if $\widetilde{T}$ and $\widetilde{Z}$ are independent conditionally on $\widetilde{A}_t$, and if the support of T is contained in that of C, then T and Z are independent conditionally on A_t.*

Proof. See the Appendix.

By Theorem 2, when $\widetilde{\tau}_t \neq 0$ for some t we may conclude that the process is not Markovian. This is interesting, because the variables $\widetilde{T}$ and $\widetilde{Z}$ in which $\widetilde{\tau}_t$ is based are completely observable, and hence one may use ordinary estimators. However, one should not take the value $\widetilde{\tau}_t$ as the (local) future-past association between the original variables of the process, since $\widetilde{\tau}_t$ and τ_t will be different in non Markovian situations. In order to illustrate this, in Figure 3 we report the traces of $\widetilde{\tau}_t$ and τ_t for Models 1 and 2, for several censoring degrees; we can appreciate how both traces become more and more distinct as the censoring grows. For the Markovian Model 0, however, both traces coincide (and they collapse to zero).

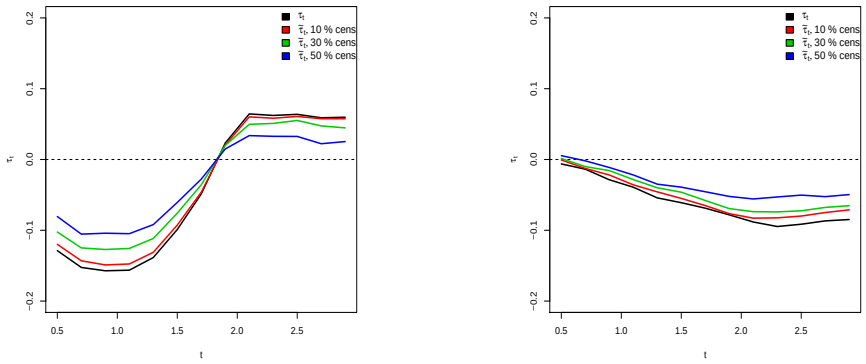

Figure 3: Comparision between τ_t and $\tilde{\tau}_t$ for several censoring rates in non-markovian simulated models. Monte Carlo approximation based on 100,000 observations. Left panel: Model 1. Right panel: Model 2.

We define $\widehat{\widetilde{\tau}}_t$ in the obvious way: $\widehat{\widetilde{\tau}}_t = \widehat{\widetilde{p}}_{c,t} - \widehat{\widetilde{p}}_{d,t}$ where

$$\widehat{\widetilde{p}}_{c,t} = 2\int\int \widehat{\widetilde{F}}_t(x^-, y^-)\widehat{F}_t(dx, dy), \qquad \widehat{\widetilde{p}}_{d,t} = 2\int\int \widehat{\widetilde{U}}_t(x, y)\widehat{\widetilde{F}}_t(dx, dy),$$

where $\widehat{\widetilde{U}}_t(x, y) = \widehat{\widetilde{F}}_t(t, y^-) - \widehat{\widetilde{F}}_t(x, y^-)$, and where $\widehat{\widetilde{F}}_t$ is the ordinary empirical distribution function of the n_t pairs $\left(\widetilde{Z}_i, \widetilde{T}_i\right)$ satisfying $\widetilde{Z}_i \leq t < \widetilde{T}_i$. This is just the ordinary Kendall's tau computed from a subsample. Results for $\widehat{\widetilde{\tau}}_t$ in Tables 4 to 6 correspond to rejection proportions of the null hypothesis $\widetilde{H}_{0t}$ when using $\left|\widehat{\widetilde{\tau}}_t\right|$ as a test statistic. Although there are other well-known procedures, in order to facilitate the comparison with $|\widehat{\tau}_t|$, critical points were obtained by bootstrapping the null hypothesis; explicitly, we used the following resampling plan:

Step 1. Draw $\widetilde{Z}_i^*$ from $\widehat{\widetilde{F}}_t(., \infty)$; that is, $P^*(\widetilde{Z}_i^* = \widetilde{Z}_i) = I(\widetilde{Z}_i \leq t < \widetilde{T}_i)/n_t$ where $n_t = \sum_{j=1}^{n} I(\widetilde{Z}_j \leq t < \widetilde{T}_j)$.

Step 2. Draw independently $\widetilde{T}_i^*$ from $\widehat{\widetilde{F}}_t(\infty, .)$; that is, $P^*(\widetilde{T}_i^* = \widetilde{T}_i) = I(\widetilde{Z}_i \leq t < \widetilde{T}_i)/n_t$.

Results for $\widehat{\widetilde{\tau}}_t$ in Tables 4 to 6 indicate that this test behaves essentially as $\widehat{\tau}_t$ but it resists much better the censoring effects. Hence, in practice, one may prefer it as a testing strategy.

In Figure 4 (bottom left panel) we give the trace of values of $\widehat{\widetilde{\tau}}_t$ for the bladder cancer data, with a 95% pointwise confidence band based on the simple bootstrap. This Figure 4 (bottom left panel) resembles Figure 4 (top left panel) corresponding to $\widehat{\tau}_t$, reaching significance about negative future-past association at similar t-values. Some differences at the right of the grid can be appreciated, where $\widehat{\widetilde{\tau}}_t$ turns up to (non-significative) positive association while $\widehat{\tau}_t$ is still negative; this is due to the influence of the censoring.

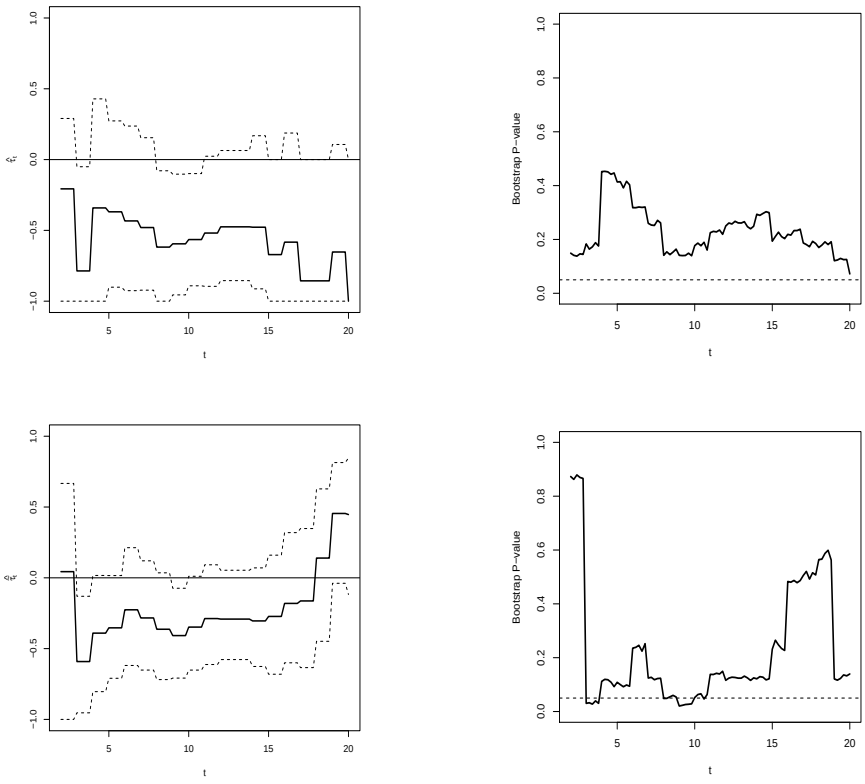

Figure 4: Left: Future-past association trace for the Bladder cancer data and pointwise confidence limits (95%). Top left panel: τ_t, bottom left panel: $\widetilde{\tau}_t$. Rigth: Significance trace. Top right panel: τ_t , bottom right panel: $\widetilde{\tau}_t$. B=2000 bootstrap resamples

Censoring also influences the width of the band, which is wider for $\widehat{\tau}_t$. Finally, in Figure 4 (top and bottom right panel) we report the significance trace of $\widehat{\widetilde{\tau}}_t$ and $\widehat{\tau}_t$ respectively, constructed from their respective bootstrap p-values

$$\frac{1}{B}\sum_{b=1}^{B} I\left(\left|\widehat{\widetilde{\tau}}_t^{*b}\right| > \left|\widehat{\widetilde{\tau}}_t\right|\right) \qquad \text{and} \qquad \frac{1}{B}\sum_{b=1}^{B} I\left(\left|\widehat{\tau}_t^{*b}\right| > \left|\widehat{\tau}_t\right|\right).$$

We see that the significance trace of $\widehat{\widetilde{\tau}}_t$ is able to detect non-markovianity (taking values below $\alpha = 0.05$ for $t \approx 3$ and $t \approx 9$) while no statistical significance is got from Figure 4 (top left panel); this is in agreement with the lower power of the test statistic $\widehat{\tau}_t$ in the simulations.

4 Conclusions and final discussion

In this paper a new method to measure future-past association for the three-state progressive model has been introduced. The proposed method is based on the Kendall's Tau association measure, adapted to the possibility of censoring, and defined in a local way. Hence, the method provides a trace of association values along time, which should be close to zero if the underlying process is Markov. Consistency and asymptotic normality of the local association measure have been established. It has been also demonstrated that, despite of possible systematic biases due to censoring effects, the proposed measure of association $\widehat{\tau}_t$ vanishes in the limit whenever the target τ_t does. Finite sample performance of the local Kendall's Tau has been assessed through simulations.

On the basis of the introduced method, a formal test for markovianity has been introduced. The test rejects the null hypothesis of markovianity when the absolute value of $\widehat{\tau}_t$ is large. A bootstrap resampling plan has been proposed to approximate the null distribution of the test statistic, and its behaviour has been evaluated through simulations. In practice, a grid of t points must be chosen so the test can be performed along time. This grid should contain a number of relevant points on the support of the lifetime of the process. After the application of the test, one comes up with a trace of p-values along the grid points; this 'significance trace' can be used to judge the (lack of) markovianity of the process along its whole evolution. Eventually, one may reach the conclusion that the process behaves in a markovian way on a certain time interval (which reports non-significative future-past association values), while getting evidence of non-markovianity on the rest of its support (where $\tau_t \neq 0$). Interestingly, a different test statistic $\widehat{\widetilde{\tau}}_t$ has been introduced, with the nice property of preserving more power than $\widehat{\tau}_t$ when the censoring level increases. All the methods have been illustrated through the analysis of real data on bladder cancer recurrence times.

Still, since many local tests are performed at the same time, one may ask about the increase of type I error rates. Formally this is a real issue, and at least

two directions could be followed to avoid it. The first one is introducing data-driven algorithms to select the 'best' t point with respect to a given criterion, e.g. maximizing the power. The second option is considering the test statistic $\widehat{\tau}_t$ (or $\widehat{\widehat{\tau}}_t$) as a process indexed by t, taking some summary measure (e.g. the supremum absolute value), and deriving its null distribution. Both possibilities are currently under investigation. However, we still believe that the introduced methods are useful since they provide a deep inspection of the process all along its evolution.

It would be interesting to extend the given methods to other more involved multi-state models. We believe that an extension to the illness-death progressive model should be more or less direct. We are presently exploring this issue. The extension to other multi-state models can become complicated as the number of states and possible transition grow. However, the method could be applied as it is to investigate markovianity in specific parts of a general multi-state model, under the assumption that the only source of possible non-markovianity is the sojourn time in the previously visited state.

Acknowledgements

Work supported by the Grants MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovación and 10PXIB300068PR of the Xunta de Galicia. Financial support from the INBIOMED project (DXPCTSUG, Ref. 2009/063) is also acknowledged.

References

[1] Andersen, P.K., Esbjerg, S. and Sorensen T.I. (2000). Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* **19**, 587–599.

[2] Andersen, P.K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research* **11**, 91–115.

[3] Byar, D. P. (1980). The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine and Topical Thiotepa, in *Bladder Tumors and Other Topics in Urological Oncology*, eds. M. Pavone-Macaluso, P.H. Smith, and F. Edsmyn, New York: Plenum, pp. 363-370.

[4] Cook, R.J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer

[5] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

[6] Kay, R. (1986). A markov model for analysing cancer markers and disease states in survival analysis. *Biometrics* **42**, 855–865.

[7] Meira-Machado, L., de Uña-Álvarez J. and Cadarso-Suárez, C. (2006). Non-parametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis* **12**, 325–344.

[8] Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C. and Andersen, P.K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* **18**, 195–222.

[9] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89–103.

[10] Stute, W. (1995). The central limit theorem under random censorship. *Annals of Statistics* **23**, 422-439.

[11] Stute, W. (1996). Distributional convergence under random censroship when covariables are present. *Scandinavian Journal of Statistics* **23**, 461–471.

[12] Wang, W. and Wells, M.T. (2000). Estimation of Kendall's Tau under censoring. *Statistics Sinica* **10**, 1199–1215.

[13] Wei, L. J., Lin, D.Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.

5 Appendix: Technical proofs

Proof to Theorem 1. The estimator

$$\widehat{F}_t(x,y) = \frac{1}{\widehat{P}(A_t)} \sum_{i=1}^{n} W_{ni} I\left(\widetilde{Z}_{i:n} \leq x, \widetilde{T}_{i:n} \leq y\right) I(\widetilde{Z}_{i:n} \leq t < \widetilde{T}_{i:n})$$

can be written as

$$\widehat{F}_t(x,y) = \frac{S_n(\varphi_{t,x,y})}{\widehat{S}_T(t) - \widehat{S}_Z(t)}$$

where $\varphi_{t,x,y}(u,v) = I(u \leq x, v \leq y)I(u \leq t < v)$ and where, for each bivariate function φ, $S_n(\varphi)$ stands for the multivariate Kaplan-Meier integral

$$S_n(\varphi) = \sum_{i=1}^{n} W_{ni} \varphi\left(\widetilde{Z}_{i:n}, \widetilde{T}_{i:n}\right) = \sum_{i=1}^{n} W_{ni} \varphi\left(Z_{i:n}, \widetilde{T}_{i:n}\right).$$

The latter equality follows because $W_{ni} = 0$ whenever $Z_{i:n}$ is censored. Now, by the Theorem in Stute (1993), we have

$$S_n(\varphi_{t,x,y}) \to E\left[\varphi_{t,x,y}(Z,T) I(T \leq b_H)\right] = P(Z \leq x, T \leq y, A_t, T \leq b_H) \qquad \text{w.p. } 1$$

where we have used continuity of H and assumptions (i) and (ii) in page 91 of the referred paper, which hold because C is independent of (Z,T). On the other hand, by using standard results for univariate Kaplan-Meier estimators, we have

$$\widehat{S}_T(t) - \widehat{S}_Z(t) \to P(Z \leq t, Z \leq b_L) - P(T \leq t, T \leq b_H) \qquad \text{w.p. 1}$$

where $b_L = \inf\{t : L(t) = 1\}$ and L is the df of $\widetilde{Z}$. Note that, if restriction $Z \leq b_L$ is not superfluous, then $b_H = b_L$; and when it is superfluous, we also may replace b_L by b_H in the above limit because (since $Z \leq T$ w.p. 1) we always have $b_L \leq b_H$. Hence,

$$\begin{aligned} P(Z &\leq t, Z \leq b_L) - P(T \leq t, T \leq b_H) = P(Z \leq t \wedge b_H) - P(T \leq t \wedge b_H) \\ &= P(Z \leq t \wedge b_H) - P(T \leq t \wedge b_H, Z \leq t \wedge b_H) \\ &= P(Z \leq t \wedge b_H < T) = P(A_{t \wedge b_H}). \end{aligned}$$

In sum, for $t \leq b_H$, w.p. 1

$$\widehat{F}_t(x,y) \to \frac{P(Z \leq x, T \leq y, A_t, T \leq b_H)}{P(A_t)} = P(Z \leq x, T \leq y, T \leq b_H \mid A_t).$$

Note that the limit is just $F_t(x,y)$ when $y \leq b_H$. Also, by applying a standard uniformity argument (cfr. Stute, 1993), we get

$$\sup_{x,y \leq b_H} \left| \widehat{F}_t(x,y) - F_t(x,y) \right| \to 0 \qquad \text{w.p. 1.}$$

Note that by applying the same arguments to

$$\begin{aligned} \int\int \psi(x,y) \widehat{F}_t(dx,dy) &= \frac{1}{\widehat{P}(A_t)} \sum_{i=1}^{n} W_{ni} \psi\left(\widetilde{Z}_{i:n}, \widetilde{T}_{i:n}\right) I(\widetilde{Z}_{i:n} \leq t < \widetilde{T}_{i:n}) \\ &= \frac{S_n(\psi_t)}{\widehat{S}_T(t) - \widehat{S}_Z(t)} \end{aligned}$$

where $\psi_t(x,y) = \psi(x,y)\, I(x \leq t < y)$, we get

$$\int\int \psi(x,y) \widehat{F}_t(dx,dy) = \int\int_{\{y \leq b_H\}} \psi(x,y) \widehat{F}_t(dx,dy) \to$$

$$\frac{E[\psi(Z,T) I(T \leq b_H) I(A_t)]}{P(A_t)} = \int\int_{\{y \leq b_H\}} \psi(x,y) F_t(dx,dy) \qquad (1)$$

w.p. 1 for each ψ function.

Now, we have

$$\widehat{p}_{c,t} = 2 \int\int \widehat{F}_t(x^-,y^-) \widehat{F}_t(dx,dy) = 2 \int\int_{\{y \leq b_H\}} \widehat{F}_t(x^-,y^-) \widehat{F}_t(dx,dy)$$

and hence

$$\begin{aligned}
\widehat{p}_{c,t} - p^0_{c,t} &= 2\int\int_{\{y\le b_H\}} \widehat{F}_t(x^-,y^-)\widehat{F}_t(dx,dy) \\
&\quad -2\int\int_{\{y\le b_H\}} F_t(x^-,y^-)F_t(dx,dy) \\
&= 2\int\int_{\{y\le b_H\}} F_t(x^-,y^-)[\widehat{F}_t(dx,dy) - F_t(dx,dy)] + \\
&\quad +2\int\int_{\{y\le b_H\}} [\widehat{F}_t(x^-,y^-) - F_t(x^-,y^-)]\widehat{F}_t(dx,dy) \\
&\equiv I + II.
\end{aligned}$$

For I we have $I \to 0$ w.p. 1 by applying the previous result (1) to the special function $\psi(x,y) = F_t(x^-,y^-)$, while for II we have

$$|II| \le \sup_{x,y\le b_H} \left|\widehat{F}_t(x,y) - F_t(x,y)\right| \int\int_{\{y\le b_H\}} \widehat{F}_t(dx,dy) \to 0 \qquad \text{w.p. 1.}$$

This concludes with the proof of $\widehat{p}_{c,t} \to p^0_{c,t}$ w.p. 1. Now, by noting

$$\sup_{x,y\le b_H} \left|\widehat{U}_t(x,y) - U_t(x,y)\right| \le 2 \sup_{x,y\le b_H} \left|\widehat{F}_t(x,y) - F_t(x,y)\right|$$

and by using result (1) similarly as before, we get $\widehat{p}_{c,t} \to p^0_{c,t}$ w.p. 1, and the proof of the consistency of $\widehat{\tau}_t$ is complete.

Now we prove the asymptotic normality of $n^{1/2}(\widehat{\tau}_t - \tau^0_t) = n^{1/2}(\widehat{p}_{c,t} - p^0_{c,t}) - n^{1/2}(\widehat{p}_{d,t} - p^0_{d,t})$. We only prove the asymptotic normality of $n^{1/2}(\widehat{p}_{c,t} - p^0_{c,t})$ due to the existing similarities between the probability of concordance and discordance. We write

$$\begin{aligned}
n^{1/2}(\widehat{p}_{c,t} - p^0_{c,t}) &= 2\int\int_{\{y\le b_H\}} \widehat{W}_t(x,y)F_t(dx,dy) \\
&\quad +2\int\int_{\{y\le b_H\}} F_t(x^-,y^-)\widehat{W}_t(dx,dy) \\
&\quad +2\int\int_{\{y\le b_H\}} \widehat{W}_t(x,y)[\widehat{F}_t(dx,dy) - F_t(dx,dy)],
\end{aligned}$$

where $\widehat{W}_t(x,y) = n^{1/2}(\widehat{F}_t(x,y) - F_t(x,y))$. We may transform $\widehat{W}_t(x,y)$ into a multivariate Kaplan-Meier process in the sense of Stute (1996) by noting that

$$n^{-1/2}\widehat{W}_t(x,y) =$$

$$\frac{1}{P(A_t)}[\sum_{i=1}^{n} W_{ni} I\left(\widetilde{Z}_{i:n} \le x, \widetilde{T}_{i:n} \le y\right) I(\widetilde{Z}_{i:n} \le t < \widetilde{T}_{i:n}) - E[I(Z \le x, T \le y)I(A_t)] +$$

$$+\left[\frac{1}{\widehat{P}(A_t)}-\frac{1}{P(A_t)}\right]\sum_{i=1}^{n}W_{ni}I\left(\widetilde{Z}_{i:n}\leq x,\widetilde{T}_{i:n}\leq y\right)I(\widetilde{Z}_{i:n}\leq t<\widetilde{T}_{i:n}),$$

that is,

$$\begin{aligned}\widehat{W}_t(x,y) &= n^{1/2}\frac{1}{P(A_t)}\left[S_n(\varphi_{t,x,y})-S(\varphi_{t,x,y})\right]+n^{1/2}S_n(\varphi_{t,x,y})\left[\frac{1}{\widehat{P}(A_t)}-\frac{1}{P(A_t)}\right]\\ &\sim n^{1/2}\frac{1}{P(A_t)}\left[S_n(\varphi_{t,x,y})-S(\varphi_{t,x,y})\right]+n^{1/2}\frac{S(\varphi_{t,x,y})}{P(A_t)^2}\left[P(A_t)-\widehat{P}(A_t)\right].\end{aligned}$$

where we put $S(\varphi)=E[\varphi(Z,T)]$. Convergence of the finite-dimensional distributions of $\widehat{W}_t(x,y)$ to a normal follows by the arguments in Stute (1996) and the multivariate CLT. Note that conditions (1.4)-(1.6) in Stute (1996) hold for $\varphi_{t,x,y}$ under $\int_0^{b_H}(1-G)^{-1}dF_T<\infty$ when $F_T(b_H)<1$ or when $F_T(b_H)=1$ and $G(b_H)<1$. Tightness also follows because of the special form of the function $\varphi_{t,x,y}$. The lengthy details are ommitted. In sum, $\widehat{W}_t(x,y)$ is weakly convergent to a Gaussian process. The rest of the proof follows the steps in Wang and Wells (2000), pp. 1212-1213.■

Proof to Theorem 2. We begin stating two useful Lemmas. Put $F_{1t}(x)=F_t(x,\infty)$ and $F_{2t}(y)=F_t(\infty,y)=F_t(t,y)$ for the marginal distributions of Z and T respectively given A_t. When needed, we also use the notations $\widetilde{F}_{1t}(x)=\widetilde{F}_t(x,\infty)=\widetilde{F}_t(x,b_H)$ and $\widetilde{F}_{2t}(y)=\widetilde{F}_t(\infty,y)=\widetilde{F}_t(t,y)$ for the marginal distributions of $\widetilde{Z}$ and $\widetilde{T}$ respectively given $\widetilde{A}_t$.

Lemma A. We have for all $x\leq t<y\leq b_H$

$$\widetilde{F}_t(x,y)=F_{1t}(x)-P(C>y|C>t)\left[F_{1t}(x)-F_t(x,y)\right].$$

Proof. Write for $x\leq t<y\leq b_H$

$$\begin{aligned}\widetilde{F}_t(x,y) &= P(\widetilde{Z}\leq x,\widetilde{T}\leq y|\widetilde{A}_t)=P(Z\leq x,T\wedge C\leq y|Z\leq t<T,C>t)\\ &= P(Z\leq x|Z\leq t<T,C>t)-P(Z\leq x,T\wedge C>y|Z\leq t<T,C>t)\\ &= F_{1t}(x)-P(Z\leq x,T>y|Z\leq t<T)P(C>y|C>t)\\ &= F_{1t}(x)-P(C>y|C>t)\left[F_{1t}(x)-F_t(x,y)\right].\blacksquare\end{aligned}$$

Corollary. We have $\widetilde{F}_{1t}(x)=F_{1t}(x)$ for all $x\leq t$, and

$$\widetilde{F}_{2t}(y)=1-P(C>y|C>t)\left[1-F_{2t}(y)\right]\qquad\text{for all } y\leq b_H.$$

Proof. For the first assertion note that, by Lemma A,

$$\widetilde{F}_{1t}(x)=\widetilde{F}_t(x,b_H)=F_{1t}(x)-P(C>b_H|C>t)\left[F_{1t}(x)-F_t(x,b_H)\right]=F_{1t}(x),$$

where the last equality follows because $P(C > b_H | C > t) = 0$ whenever $F_t(x, b_H) < F_{1t}(x)$. The second assertion follows directly from Lemma A.■

Lemma B. The two following conditions are equivalent:

(i) $\widetilde{F}_t(x, y) = \widetilde{F}_{1t}(x)\widetilde{F}_{2t}(y)$ for all $x \leq t < y \leq b_H$ (i.e. $\widetilde{Z}$ and $\widetilde{T}$ are independent given $\widetilde{A}_t$)

(ii) $F_t(x, y) = F_{1t}(x)F_{2t}(y)$ for all $x \leq t < y \leq b_H$

Proof. Assume that (i) holds. Then, by Lemma A and its Corollary, we have

$$\begin{aligned} F_{1t}(x) - P(C &> y | C > t) \left[F_{1t}(x) - F_t(x, y)\right] = \\ &= F_{1t}(x) - F_{1t}(x) P(C > y | C > t) \left[1 - F_{2t}(y)\right], \end{aligned}$$

which holds only if

$$P(C > y | C > t) F_t(x, y) = P(C > y | C > t) F_{1t}(x) F_{2t}(y).$$

But this is just (ii), after noting that $P(C > y | C > t) > 0$ for $y < b_H$. Conversely, if (ii) holds, then (i) inmediately follows from Lemma A and its Corollary.■

Lemma B implies the two assertions of Theorem 2. Note that, when the support of T is contained in that of C, we have that (ii) holds if and only if Z and T are conditionally independent given A_t.■