Universidade de Vigo

A nonparametric test for markovianity in the illness-death model

Mar Rodríguez Girondo and Jacobo de Uña Álvarez

Report 11/04

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://webs.uvigo.es/depc05/
E-mail: depc05@uvigo.es

Universidade de Vigo

A nonparametric test for markovianity in the illness-death model

Mar Rodríguez Girondo and Jacobo de Uña Álvarez

Report 11/04

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

A nonparametric test for markovianity in the illness-death model

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

[1]SiDOR Research Group, University of Vigo. Address: Facultade de CC Económicas e Empresariais, Campus Lagoas-Marcosende, 36310, Vigo, Spain. E-mail: margirondo@uvigo.es

[2]Department of Statistics and OR, University of Vigo. Address: Facultade de CC Económicas e Empresariais, Campus Lagoas-Marcosende, 36310, Vigo, Spain. E-mail: jacobo@uvigo.es

Abstract

Multi-state models are useful tools for modeling disease progression when survival is the main outcome but several intermediate events of interest are observed during the follow-up time. The illness-death model is a special multi-state model with important applications in the biomedical literature. It provides a suitable representation of the individual's history when an unique intermediate event can be experienced before the main event of interest. Nonparametric estimation of transition probabilities in this and other multi-state models is usually performed through the Aalen-Johansen estimator under a Markov assumption . The Markov assumption claims that given the present state, the future evolution of the illness is independent of the states previously visited and the transition times among them. However, this assumption fails in some applications, leading to inconsistent estimates. In this paper we provide a new approach for testing markovianity in the illness-death 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. Besides, we propose a global test for markovianity based on a supremum-type test statistic. The finite sample performance of the test is investigated through simulations. We illustrate the new method through the analysis of two biomedical data analysis.

Keywords Multi-state models; Illness-death model, Markov condition; Kendall's Tau.

1 Introduction

Multi-state models [1, 2, 3] are typically used for modeling disease progression when several intermediate events of interest are observed during the follow-up time. These models are an extension of the traditional survival analysis and make it possible to account for complex individuals' history with a possible influence on the prognosis. In particular, the traditional survival analysis refers to the simplest multi-state model, the mortality model, where only two states are considered, an initial ('alive') state and a final absorbing state ('dead').

Mathematically, a multi-state model refers to a stochastic process varying in continuous time $\{X(t), t \geq 0\}$, where $X(t)$ is the state of the individual at time t and allowing individuals to move along a finite number of states. Furthermore, we assume that the trajectories of individuals can be right censored by a potential censoring time that is independent of the process. 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 HIV infection), or a non-fatal complication in the course of the illness (e.g. cancer recurrence).

In this work, we focus on the illness-death model, which involves three different states by splitting the 'alive' state in two different states (1='healthy', 2='dis-

eased') and considering a last absorbing state (3='dead') and three possible transitions among them $1 \rightarrow 2$, $2 \rightarrow 3$ and $1 \rightarrow 3$ (see Figure 1). Many applications of the illness-death model can be found in the biomedical literature [4, 5, 6, 7, 8]. In section 4, we reanalyze two biomedical datasets in a new scope. Firstly, we propose to consider an illness-death model for analyzing the influence of the time to chronic graft-versus-disease in evolution after a bone marrow transplant in patients with leukemia [9]. Secondly, we consider the relationship between the time to the first wound excision and time to Straphylocous Aureaus infection in burn patients [10]. It is noteworthy that the illness-death model contains other simpler schemes, such as the mortality model and the three-state progressive model where direct transitions between state 1 and state 3 are not possible.

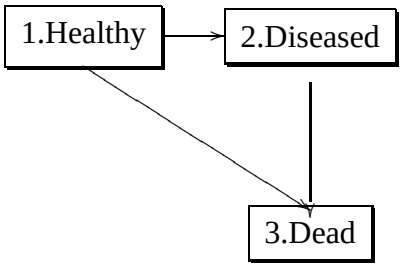

Figure 1: Illness-death model

Denoting by Z the sojourn time in state 1 and by T the total survival time, the illness-death model is characterized by the joint distribution of (Z, T). Note that individuals who transit directly from state 1 to state 3 are those with $Z = T$, while $Z < T$ indicates that the individual has passed through the intermediate state 2. The Markov assumption claims that given the present state, the future evolution of the illness is independent of the states previously visited and the transition times among them. This condition is sometimes violated. For example, [6] 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. One of the main arguments for checking the Markov condition is that the usual estima-

tors (e.g. Aalen-Johansen transition probabilities) may be systematically biased in non-Markovian situations [11]. Despite non-Markovian estimators for transition probabilities and related curves are available [11, 12, 13], it has been quoted that including the Markov information in the construction of the estimators allows for variance reduction [11]. So, in practice, one will be interested in testing markovianity.

Let $\lambda_s(.)$ denote the hazard function of T for those individuals going from state 1 to 2 exactly at time s (that is, for the subpopulation $\{Z = s, Z < T\}$). Note that for $t \geq s$, the hazard $\lambda_s(t)$ is the instantaneous probability of 'death' for those individuals being in state 2 at time t, which may be influenced by the specific transition time s. The Markov assumption states that $\lambda_s(t)$ does not depend on the particular value of s, $t \geq s$. Traditionally, the Markov assumption is checked by testing $H_0 : \beta = 0$ under the proportional hazard model [4] $\lambda_s(t) = \lambda_0(t)e^{\beta s}$, $t \geq s$. However, the proportional hazard specification may not properly represent the dependence structure between the survival prognosis and the sojourn time in state 1 since both the linearity and proportional hazards may fail in practice. As a consequence, this approach may be unable to detect the lack of markovianity. See also sections 3 and 4. Therefore, alternative, more flexible methods are required.

Alternatively, the Markov condition can be formulated as the independence between variables Z and T conditionally on the event $A_t = \{Z \leq t < T\}$ ('being in state 2 at time t'), for each given $t > 0$. If we denote by $H_{0t} : T \perp Z$ given A_t, $t > 0$, where $\perp$ stands for the independence relationship, then $H_0 = \bigcap_t H_{0t}$ represents the Markov condition in the illness-death model framework. The general idea of the new methods is based on determining -without assuming any predefined structure- the grade of dependency between past (Z) and the future (T) given the present. For this, the Kendall's tau will be used in a local way.

The rest of the paper is organized as follows. In Section 2 we introduce the new goodness-of-fit method for testing markovianity in the illness-death model based on measuring future-past association. We introduce a bootstrap resampling

plan to approximate the null distribution of the test statistic. In section 3 a simulation study shows the performance of the new methods. Also, in section 4 the new methods are applied to two real biomedical datasets and compared to the traditional Cox proportional hazards approach. Main conclusions and a final discussion follow in section 5, while the technical details are deferred to the Appendix.

2 Main results

2.1 Goodness-of-fit tests based on Kendall's tau

Focusing on the illness-death model discussed in the Introduction, let Z and T be the sojourn time in state 1 and the time to reach state 3, respectively. As indicated, 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 $U_t(x, y) = P(Z > x, T < y | A_t) = F_t(t, y^-) - F_t(x, y^-)$.

We firstly propose a method to test the null hypothesis $H_{0t} : \tau_t = 0$ for each fixed time point. The basic idea is to reject H_{0t} for large values of an estimator of $|\tau_t|$. Since the Markov condition is represented by the intersection null $H_0 = \bigcap_{t>0} H_{0t}$, evidence against any specific H_{0t} can be interpreted as a lack of

markovianity of the process. On the other hand, since many local tests are performed at the same time, one may ask about the increase of type I error rates due to the multiplicity of local tests . As a solution, we propose a global test taking $\sup_t |\tau_t|$ as a summary measure.

In most practical cases, 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).

Therefore, for censored samples, we can introduce an alternative association measure, $\widetilde{\tau}_t$, which is defined as the 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 $t > 0$; $\widetilde{\tau}_t = \widetilde{p}_{c,t} - \widetilde{p}_{d,t}$ where

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

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 $\widetilde{H}_{0t} : \widetilde{\tau}_t = 0$. Interestingly, when T and Z are independent conditionally on A_t we have that $\widetilde{Z}$ and $\widetilde{T}$ are independent given $\widetilde{A}_t = \left\{\widetilde{Z} \leq t < \widetilde{T}\right\}$. The converse is also true

provided that the support of C contains that of T . This is stated as a Theorem.

Theorem 1 *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.

Theorem 1 is useful, because it ensures that if $\widetilde{\tau}_t \neq 0$ for some t we may conclude that the process is not Markovian. Since variables $\widetilde{T}$ and $\widetilde{Z}$ in which $\widetilde{\tau}_t$ is based are completely observable, one may use ordinary estimators. This avoids the use of Kaplan-Meier-based estimators, which typically exhibits a large variance under heavy censoring, and hence they lead to tests with a small statistical power. See [14] for more on this in the scope of the three-state progressive model.

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 (except for the uncensored case, in which $\widetilde{\tau}_t = \tau_t$). In order to illustrate this, in Figure 2 we report the traces of $\widetilde{\tau}_t$ and τ_t for the non-Markovian Model 1 in section 3, for several censoring degrees; we can appreciate how both traces become more and more distinct as the censoring grows. Indeed, this Figure indicates that much less power should be expected under heavy censoring, because the trace of the Kendall's tau is closer to zero. For the Markovian Model 0 in section 3, however, both traces coincide (and they collapse to zero).

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

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

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 $\widetilde{Z}_i \leq t < \widetilde{T}_i$. Since, this is just the ordinary Kendall's tau computed from a subsample, results on consistency, finite sample distribution, and asymptotic normality are well-known. We refer the

interested reader to [15] and [16]. Hence, critical points for the local test under the null can be computed as usual.

Problems arise when considering the statistic for the global test, namely $D_n = \sup_t \left|\widehat{\widetilde{\tau}}_t\right|$. The asymptotic normality of the finite-dimensional distributions of the process $(\widehat{\widetilde{\tau}}_t)_{t>0}$ follows from the standard results on the Kendall's tau and the multivariate Central Limit Theorem. Therefore, the weak convergence of $\left(\widehat{\widetilde{\tau}}_t\right)_t$ to a Gaussian process -and hence the existence of a limiting distribution for D_n -follows provided that the process is tight. This is stated by Theorem 2 below. However, the computation of the critical points from the asymptotic distribution of D_n is difficult, and therefore we propose in the following subsection a bootstrap resampling plan.

Theorem 2. Let $\mathcal{I}$ be an interval such that $P(\widetilde{A}_t) \geq c > 0$ for all $t \in \mathcal{I}$. Then, the process $\left\{\sqrt{n}(\widehat{\widetilde{\tau}}_t - \widetilde{\tau}_t) : t \in \mathcal{I}\right\}$ weakly converges to a zero-mean Gaussian process.

Proof: See the Appendix.

2.2 Approximation of null distribution. Bootstrap approaches

In this subsection we propose a bootstrap approximation to the null distribution of both $\widehat{\widetilde{\tau}}_t$ and $D_n = \sup_t \left|\widehat{\widetilde{\tau}}_t\right|$. Although this is not needed for the local test (due to the availability of tables with critical points and asymptotic theory), for completeness we give a bootstrap proposal (termed as Local Bootstrap, LB) also for this case. The LB draws independently the $\widetilde{Z}$ and the $\widetilde{T}$ at each $\widetilde{A}_t$. Therefore, it is specific for each particular t value. On the other hand, to deal with the global test a Global Bootstrap (GB) resampling plan is needed. This is also given below. Note that the GB may also be applied to the local test, in order to incorporate the full Markov information in the (local) testing procedure.

Local Bootstrap (LB):

The local bootstrap resample plan proceeds in two steps:

Step 1: Draw $\widetilde{Z}_i^*$ from the marginal distribution $\widehat{\widetilde{F}}_t(.,\infty)$.
Step 2: Draw independently $\widetilde{T}_i^*$ from the marginal distribution $\widehat{\widetilde{F}}_t(\infty,.)$.

In Step 1, each $\widetilde{Z}_i$ corresponding to a pair $(\widetilde{Z}_i, \widetilde{T}_i)$ falling on $\widetilde{A}_t$ is sampled (with replacement) with probability $1/n_t$, where $n_t = \sum_{j=1}^{n} I(\widetilde{Z}_j \leq t < \widetilde{T}_j)$. In Step 2, the value $\widetilde{T}_i$ is sampled in a similar way. Note that, in this manner, new combinations of type $(\widetilde{Z}_i, \widetilde{T}_j)$ may appear in the bootstrap resample. For each fixed value of t, steps 1 and 2 are repeated until a bootstrap resample $\{(\widetilde{Z}_i^*, \widetilde{T}_i^*)\}_{i=1}^{n_t}$ of size n_t is generated. This bootstrap approach has the advantage of its simplicity and computational efficiency but it presents the drawback that it does not permit to manage the process in a global way and hence it cannot be used for the global test.

Global Bootstrap (GB):

The global bootstrap algorithm is based in the global relationship between $\widetilde{Z}$ and $\widetilde{T}$ under $\widetilde{H}_0 = \bigcap_{t>0} \widetilde{H}_{0t}$ Note that, under the null hypothesis, the process is Markov, and hence a markovian estimator of $F_{\widetilde{T}/\widetilde{Z}}(t/x) = P(\widetilde{T} \leq t | \widetilde{Z} = x, \widetilde{Z} < \widetilde{T})$ should be used to resample the total survival time given the observed transition time from 1 to 2 ($\widetilde{Z} = x$).

Explicitly, we used the following resampling algorithm:
Step 1: Draw $\widetilde{Z}_i^*$ from $\widehat{F}^1_{\widetilde{Z}}$.
Step 2: Given $\widetilde{Z}_i^*$, draw $\widetilde{T}_i^*$ from $\widehat{F}_{\widetilde{T}/\widetilde{Z}}(.|\widetilde{Z}_i^*)$

In Step 1 $\widehat{F}^1_{\widetilde{Z}}$ stands for the ordinary empirical df of $\widetilde{Z}$ given $\widetilde{Z} < \widetilde{T}$.

In Step 2, we take $\widehat{F}_{\widetilde{T}/\widetilde{Z}}(t|x) = 1 - \hat{p}_{22}(x,t)$ where

$$\hat{p}_{22}(x,t) = \prod_{x<\widetilde{T}_i \leq t, \widetilde{Z}_i < \widetilde{T}_i} \left[1 - \frac{1}{\sum_{j=1}^{n} I(\widetilde{Z}_j < \widetilde{T}_i \leq \widetilde{T}_j)} \right]$$

is the Aalen-Johansen estimator of the transition probability $p_{22}(x,t)$, which is an efficient estimator under the Markov condition.

The procedure is repeated until a bootstrap resample $\{(\widetilde{Z}_i^*, \widetilde{T}_i^*)\}_{i=1}^{n_1}$ of size $n_1 = \sum_{j=1}^{n} I(\widetilde{Z}_j < \widetilde{T}_j)$ is obtained.

3 Simulation study

We carried out a small simulation study to investigate the performance of the local test $\widehat{\widetilde{\tau}}_t$ for a fixed grid of t-points and for the global test $D_n = \sup \left|\widehat{\widetilde{\tau}}_t\right|$. Specifically, we have simulated 500 Monte Carlo trials of two different models. Both models were based in the accelerated failure time specification $\log(T - Z) = f(Z) + \varepsilon$ for the individuals passing through state 2, where the error ε is independent of the 'covariate' Z, while $f(.)$ is the predictor. An alternative representation is given by $\lambda_s(t) = \lambda_0\left((t-s)\, e^{-f(s)}\right) e^{-f(s)}$ where (recall) $\lambda_s(t)$ is the conditional hazard of T given $Z = s, Z < T$ 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 $\lambda_s(t)$ 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 $f(s) = (s-1)^2$

The traces of $\tilde{\tau}_t$ for Model 1 along the interval $t \in [0.5, 3]$ for several censoring degrees are displayed in Figure 2. From this Figure 2 we can see that Model 1 presents first a negative and then a positive future-past association, and that $\tilde{\tau}_t$ vanish for $t \approx 1.8$. This is a consequence of the increasing-decreasing shape of the conditional hazard $\lambda_s(t) = \exp\left[-(s-1)^2\right]$ on the interval $[0,2]$ (note that larger hazard values correspond to smaller T's). Of course, we omit the trace corresponding to Model 0 because in this case $\tilde{\tau}_t = 0$ for each t-value.

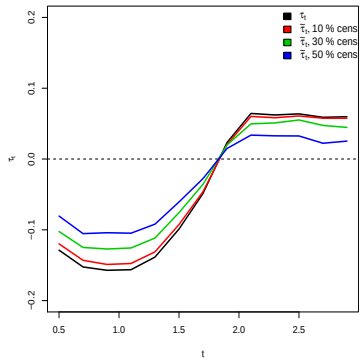

Figure 2: Model 1. Comparison between τ_t and $\widetilde{\tau}_t$ for several censoring rates in the non-markovian simulated model. Monte Carlo approximation based on 100,000 observations.

The performance of $\widehat{\widetilde{\tau}}_t$ was evaluated along a grid of t-values in the interval [0.5,3]. Specifically, we took $t = 0.7, 1.1, 1.5, 1.9, 2.3, 2.7$. The performance of $D_n = \sup|\widehat{\widetilde{\tau}}_t|$ was calculated over the same interval [0.5,3]. Two different situations in terms of direct transitions between state 1 and state 3 ($Z = T$) were considered. Firstly, we assumed a situation with no direct transitions (the particular 3-state progressive model) and a more general situation where 50% of the individuals were assumed to transit directly from state 1 to state 3. Furthermore, different sample sizes and censoring percentages were considered. As sample sizes we took $n = 100$, 250, and 500. 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. We considered four situations from the uncensored case to a maximum of 50% of censored observations.

In Tables 1 to 4 we report the rejection proportion of the local test (when based on the LB and the GB) along a grid of t-values, and for the corresponding global test (based on the GB) for Models 0 and 1. A significance level of 5% was assumed, and we took $B = 200$ bootstrap resamples. For comparison purposes, the rejection proportion corresponding to the simple method based on

a proportional hazard specification with linear predictor, $\lambda_s(t) = \lambda_0(t)e^{\beta s}$, was also included. It is interesting to see how the power of this test may not increase with the sample size, as it happens in Tables 3 and 4 (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.

Note that Model 0 is Markovian, so we expect rejection proportions about 0.05 in this case. Results in Tables 1 (no direct transitions between state 1 and state 3) and 2 (50% of direct transitions between state 1 and state 3) are quite satisfactory to this regard. The nominal level is well approximated for both bootstrap approaches in the local test. However, we observe that the the GB reports more conservative results when increasing the censoring level. The global test also presents satisfactory results, reporting rejection proportions of about 5%.

For Model 1 we expect a rejection proportion increasing with the sample size and also with the proportion of uncensoring. These features are appreciated in Tables 3 (no direct transitions between state 1 and state 3) and 4 (50% of direct transitions between state 1 and state 3). Interestingly, we see that the maximum power of the local test based on $\widehat{\widetilde{\tau}}_t$ is achieved at some central point in the grid ($t = 1.1$ in Model 1). This is a consequence of two facts. First, the amount of information on $\widetilde{\tau}_t$ grows at central points; second, we have that the largest absolute value of $\widetilde{\tau}_t$ (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). It is also remarkable the loss 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.

As for the global test , it is seen from tables 3 and 4 that its power may be much larger than that of the proportional hazards approach, particularly for low to

moderate censoring degrees. For example, with $n = 500$ and 30 % of censoring, the power of the nonparametric global test was more than four times (Table 3) or three times (Table 4) that of the proportional hazards method. However, we must indicate the loss of power observed when censoring level reaches the 50%.

Interestingly, if we compare Tables 3 and 4, we can observe a loss of power of the new methods when the number of individuals passing through the intermediate state declines. This is due to the reduction in the effective sample size to compute $\widetilde{\tau}_t$ as long as we enlarge the number of direct transitions $1 \rightarrow 3$. While in the three-state progressive model, where the direct transitions to state 3 without passing through state 2 are not allowed, all the observations are used to compute the test (Table 3), if we assume a 50% of direct transitions $1 \rightarrow 3$ (Table 4) the effective sample size is reduced to about half of the original sample size. So in practice, the proportion of observed transitions through state 2 will dramatically affect the performing of the new methods.

Table 1: Model 0 (Markovian). No transitions $1 \rightarrow 3$. Rejection proportions of tests based on $\hat{\tilde{\tau}}_t$ for several values of t and $D_n = \sup|(\hat{\tilde{\tau}}_t|)$ along 500 trials, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

% of censoring	n	Bootstrap method	0.7	1.1	1.5	1.9	2.3	2.7	D_n	PH method
	100	LB	0.042	0.050	0.028	0.062	0.040	0.062		
		GB	0.050	0.054	0.036	0.064	0.038	0.054	0.016	0.053
0	250	LB	0.040	0.049	0.045	0.046	0.043	0.051		
		GB	0.056	0.060	0.044	0.060	0.042	0.062	0.040	0.048
	500	LB	0.045	0.050	0.042	0.068	0.046	0.040		
		GB	0.046	0.064	0.058	0.076	0.044	0.042	0.044	0.060
	100	LB	0.048	0.052	0.036	0.046	0.050	0.044		
		GB	0.052	0.042	0.030	0.052	0.042	0.024	0.028	0.053
10 ($b_G = 20$)	250	LB	0.049	0.056	0.048	0.053	0.045	0.044		
		GB	0.044	0.052	0.042	0.050	0.036	0.046	0.058	0.045
	500	LB	0.041	0.042	0.049	0.063	0.053	0.042		
		GB	0.034	0.044	0.044	0.070	0.040	0.028	0.042	0.056
	100	LB	0.044	0.046	0.044	0.064	0.068	0.062		
		GB	0.020	0.026	0.028	0.042	0.052	0.020	0.032	0.053
30 ($b_G = 6.6$)	250	LB	0.045	0.046	0.049	0.058	0.049	0.053		
		GB	0.026	0.028	0.030	0.052	0.034	0.056	0.066	0.058
	500	LB	0.053	0.055	0.045	0.060	0.042	0.045		
		GB	0.030	0.032	0.040	0.038	0.030	0.024	0.050	0.058
	100	LB	0.020	0.022	0.016	0.030	0.014	0.076		
		GB	0.016	0.016	0.018	0.022	0.018	0.002	0.030	0.050
50 ($b_G = 3.9$)	250	LB	0.043	0.042	0.041	0.055	0.049	0.043		
		GB	0.014	0.020	0.020	0.020	0.022	0.026	0.080	0.048
	500	LB	0.061	0.045	0.060	0.052	0.043	0.049		
		GB	0.016	0.020	0.022	0.030	0.016	0.030	0.064	0.047

Table 2: Model 0 (Markovian). 50% transitions $1 \rightarrow 3$. Rejection proportions of tests based on $\hat{\tilde{\tau}}_t$ for several values of t and $D_n = \sup|(\hat{\tilde{\tau}}_t|)$ along 500 trials, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

% of censoring	n	Bootstrap method	0.7	1.1	1.5	1.9	2.3	2.7	D_n	PH method
	100	LB	0.064	0.064	0.054	0.062	0.066	0.084		
		GB	0.054	0.062	0.060	0.060	0.066	0.046	0.078	0.053
0	250	LB	0.046	0.046	0.034	0.048	0.055	0.042		
		GB	0.048	0.036	0.048	0.076	0.054	0.066	0.052	0.059
	500	LB	0.043	0.043	0.051	0.062	0.054	0.070		
		GB	0.038	0.062	0.036	0.068	0.040	0.012	0.046	0.057
	100	LB	0.048	0.058	0.042	0.054	0.066	0.074		
		GB	0.036	0.062	0.046	0.062	0.046	0.030	0.092	0.053
8 ($b_G = 20$)	250	LB	0.051	0.048	0.049	0.049	0.044	0.049		
		GB	0.038	0.064	0.046	0.060	0.040	0.040	0.048	0.062
	500	LB	0.063	0.047	0.050	0.049	0.051	0.068		
		GB	0.056	0.056	0.040	0.060	0.052	0.078	0.042	0.063
	100	LB	0.052	0.064	0.058	0.064	0.082	-		
		GB	0.056	0.058	0.044	0.060	0.054	0.022	0.064	0.053
23 ($b_G = 6.6$)	250	LB	0.051	0.066	0.053	0.058	0.052	0.057		
		GB	0.052	0.068	0.050	0.054	0.056	0.052	0.026	0.067
	500	LB	0.050	0.049	0.060	0.061	0.046	0.069		
		GB	0.042	0.046	0.068	0.066	0.046	0.058	0.046	0.049
	100	LB	0.036	0.038	0.034	0.022	0.070	-		
		GB	0.052	0.070	0.058	0.048	0.034	0.040	0.126	0.050
36 ($b_G = 3.9$)	250	LB	0.050	0.057	0.046	0.049	0.050	-		
		GB	0.044	0.064	0.034	0.044	0.042	0.026	0.066	0.060
	500	LB	0.052	0.055	0.053	0.045	0.055	-		
		GB	0.040	0.056	0.052	0.056	0.032	0.036	0.042	0.054

Table 3: Model 1 (Non-Markovian). No transitions $1 \rightarrow 3$. Rejection proportions of tests based on $\hat{\tilde{\tau}}_t$ for several values of t and $D_n = \sup |(\hat{\tilde{\tau}}_t|)$ along 500 trials, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

% of censoring	n	Bootstrap method	0.7	1.1	1.5	1.9	2.3	2.7	D_n	PH method
	100	LB	0.142	0.165	0.120	0.062	0.074	0.090		
		GB	0.252	0.280	0.158	0.084	0.088	0.084	0.230	0.082
0	250	LB	0.448	0.610	0.369	0.082	0.177	0.144		
		GB	0.460	0.610	0.390	0.088	0.164	0.140	0.600	0.088
	500	LB	0.733	0.886	0.622	0.096	0.294	0.245		
		GB	0.736	0.884	0.616	0.116	0.292	0.258	0.882	0.087
	100	LB	0.136	0.136	0.112	0.042	0.066	0.070		
		GB	0.178	0.224	0.138	0.058	0.088	0.066	0.190	0.072
10 ($b_G = 25$)	250	LB	0.393	0.541	0.338	0.079	0.149	0.124		
		GB	0.384	0.536	0.324	0.076	0.132	0.106	0.518	0.060
	500	LB	0.665	0.845	0.526	0.092	0.238	0.189		
		GB	0.652	0.808	0.506	0.104	0.248	0.202	0.834	0.098
	100	LB	0.114	0.098	0.072	0.048	0.060	0.076		
		GB	0.086	0.162	0.082	0.042	0.056	0.044	0.104	0.072
30 ($b_G = 8.1$)	250	LB	0.277	0.403	0.222	0.068	0.099	0.086		
		GB	0.228	0.322	0.178	0.056	0.074	0.060	0.298	0.073
	500	LB	0.505	0.680	0.380	0.079	0.165	0.117		
		GB	0.442	0.618	0.368	0.068	0.146	0.100	0.482	0.118
	100	LB	0.096	0.084	0.072	0.062	0.088	-		
		GB	0.062	0.088	0.056	0.046	0.030	0.016	0.048	0.070
50 ($b_G = 4.6$)	250	LB	0.200	0.264	0.144	0.058	0.067	0.068		
		GB	0.088	0.170	0.078	0.050	0.028	0.038	0.068	0.108
	500	LB	0.371	0.462	0.239	0.069	0.102	0.076		
		GB	0.264	0.356	0.180	0.052	0.058	0.046	0.176	0.172

Table 4: Model 1 (Non-Markovian). 50% transitions $1 \rightarrow 3$. Rejection proportions of tests based on $\widehat{\widetilde{\tau}}_t$ for several values of t and $D_n = \sup |(\widehat{\widetilde{\tau}}_t|)$ along 500 trials, censoring rates and sample sizes (n). Results corresponding to the proportional hazard specification are also provided (PH method)

% of censoring	n	Bootstrap method	0.7	1.1	1.5	1.9	2.3	2.7	D_n	PH method
	100	LB	0.142	0.160	0.120	0.062	0.074	0.090		
		GB	0.140	0.184	0.116	0.052	0.078	0.082	0.102	0.082
0	250	LB	0.218	0.324	0.210	0.083	0.146	0.128		
		GB	0.222	0.326	0.200	0.096	0.150	0.134	0.275	0.089
	500	LB	0.454	0.585	0.380	0.088	0.177	0.146		
		GB	0.470	0.600	0.398	0.090	0.174	0.162	0.624	0.091
	100	LB	0.136	0.136	0.112	0.042	0.066	0.070		
		GB	0.120	0.130	0.116	0.046	0.064	0.062	0.094	0.072
8 ($b_G = 25$)	250	LB	0.202	0.284	0.164	0.081	0.123	0.110		
		GB	0.192	0.290	0.158	0.074	0.128	0.106	0.230	0.091
	500	LB	0.400	0.525	0.339	0.089	0.155	0.124		
		GB	0.406	0.534	0.366	0.092	0.168	0.136	0.522	0.093
	100	LB	0.114	0.098	0.072	0.048	0.060	0.076		
		GB	0.114	0.104	0.086	0.054	0.052	0.060	0.064	0.072
23 ($b_G = 8.1$)	250	LB	0.154	0.192	0.127	0.063	0.101	0.089		
		GB	0.156	0.178	0.108	0.070	0.108	0.076	0.142	0.067
	500	LB	0.290	0.385	0.219	0.081	0.116	0.090		
		GB	0.280	0.392	0.218	0.086	0.096	0.096	0.296	0.086
	100	LB	0.096	0.084	0.072	0.062	0.055	-		
		GB	0.088	0.086	0.072	0.056	0.050	0.034	0.256	0.070
36 ($b_G = 4.6$)	250	LB	0.094	0.143	0.080	0.046	0.071	0.068		
		GB	0.096	0.130	0.082	0.048	0.060	0.054	0.054	0.076
	500	LB	0.197	0.259	0.124	0.067	0.080	0.076		
		GB	0.206	0.292	0.132	0.078	0.076	0.060	0.090	0.098

4 Real data illustration

Two datasets were analyzed in order to illustrate the new methods. Because of the presence of ties in the real data, the discrete version of $\widehat{\widetilde{\tau}}_t$, $\overline{\tau}_t$, was used:

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

4.1 Bone marrow transplantation for leukemia

We analyzed survival data from a multicenter trial of patients with acute myeloctic leukemia after a bone marrow transplantation. The details of this dataset can be found in [9]. The recovering from a bone marrow transplant is a complex process in patients with leukemia and their prognosis may be affected by intermediate events, such as the development of chronic graft-versus-host disease (cGVHD). Therefore, we considered an illness-model to analyze the prognostic of the patients after the bone marrow transplant, considering the development of cGVHD as intermediate event of interest.

Specifically, we studied the recovery process of 136 patients. 60 of them developed cGHVD after the transplant (18% of censoring in state 1) and among them, 28 died (24% of censoring in state 2). Furthermore, 52 patients (38%) were observed to die without developing cGHVD. We studied the markovianity in the range of observed values from 150 to 500 days of follow-up time, so that we guarantee a minimum of 35 observations to perform each $\overline{\tau}_t$. In Figure 3 we report two graphics to summarize the application of the local markovianity test to this dataset. In the left panel we give the trace of values of $\overline{\tau}_t$, with its corresponding 95% and 90% pointwise confidence bands based on the simple bootstrap. This Figure suggests a significant positive future-past association in the whole range of time, that should be interpreted as an increased risk of death for those patients suffering cGVHD shortly after the bone marrow transplant. In the right panel, we report the significance trace of the local goodness-of-fit based on $\overline{\tau}_t$. Results for both local and global bootstrap methods introduced in section 2.1 are shown, and they basically report the same trace. We observe that the significance trace is able to detect non-markovianity along all the range of t-values. Accordingly, when we apply the summary test based on the supremum over the observations falling in the interval of interest, a p-value=0.014 was obtained. Interestingly, this agrees with the analysis of markovianity based on the proportional hazards model $\lambda_s(t) = \lambda_0(t)e^{\beta s}$ under which $\widehat{\beta} = -0.005$ ($s.e. = 0.00203$) with a logrank test's

p-value of 0.011.

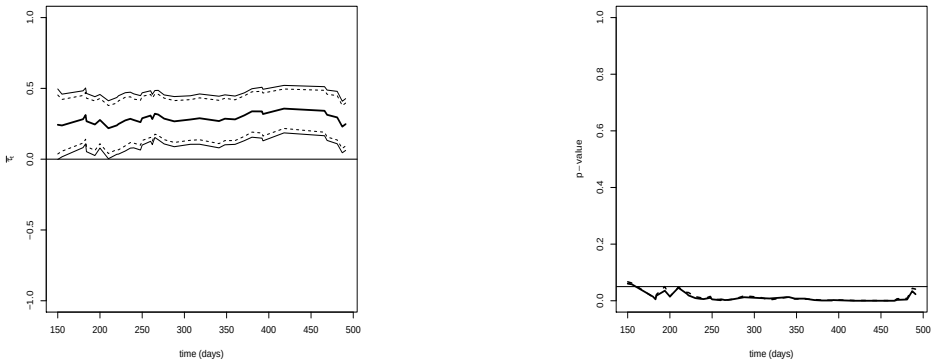

Figure 3: Left: Future-past association trace for the leukemia data and pointwise confidence limits. Solid line: 95%. Dashed line: 90%. Right: Significance trace. B=2000 bootstrap resamples. Solid line: GB. Dashed line: LB

4.2 Infection for burn patients

The second dataset refers to the study of time from admission until either infection with Staphylococcus aureus or discharge in patients from the burn unit at a large university hospital ([10]). The infection of a burn wound is a common complication for these patients that might be affected for other factors, such as days until first wound excision during the evolution process. Both, the time to Straphylocous Aureaus (T) infection and time to excision (Z) are considered in this case, and hence an illness-death model applies. From a total of 154 patients studied, the patient's wound had been excised for 84 (23% of censoring in state 1). From them, 14 patients experienced infection (83% of censoring in state 2). Besides, 34 patients experienced the infection without having passed through the intermediate event.

In this case, the Cox method accepts the global markovianity at 5% signifi-

cance level, providing a large p-value (p-value=0.448). However, when applying our methods, the results does not show that clear markovian performance. We studied the local markovianity along the interval [7,31] days from the beginning of the follow-up time to guarantee, as in the previous case, 35 observations in each corresponding $\widetilde{A}_t$. In Figure 4 (left panel), we can observe the estimated future-past association for the burn data. It suggests a negative association between the time to excision and the total time at the t-interval [14,20] days of follow-up time, meaning that the risk of Straphylocous Aureaus is increased after the excision. This is confirmed by the significance trace of the local test (Figure 4 right). The local method rejects the local markovianity at a 5% of significance level at similar part of the grid of t-points. As for the global test, we obtain p-value=0.144. Even if the global test does not reach the significance level, the p-value associated is quite lower than the one provided by the classical proportional hazard method.

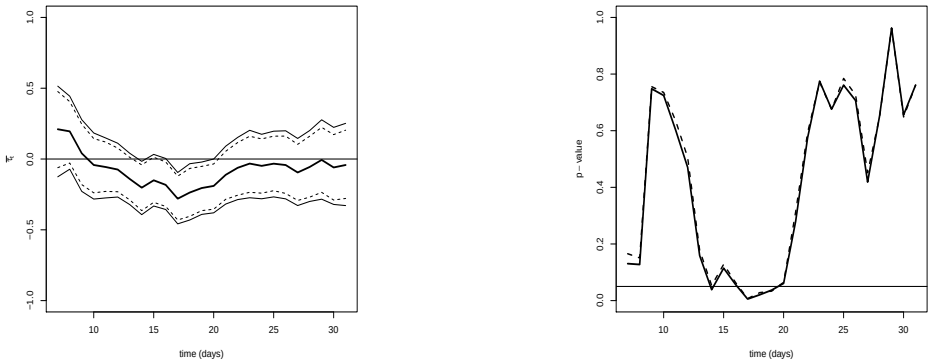

Figure 4: Left: Future-past association trace for the burn data and pointwise confidence limits. Solid line: 95%. Dashed line: 90%. Right: Significance trace. B=2000 bootstrap resamples. Solid line: GB. Dashed line: LB

5 Conclusions

In this paper, a flexible nonparametric method to test the Markov condition in the scope of the illness-death model has been proposed. The nonparametric test is based on the measured association between the past and the future of the process of interest given its present state at time t. By considering several values of t, one may construct a significance trace which reports information on the markovianity of the process in a local way. Besides, we have proposed a supremum-type test statistic by considering the maximum observed absolute future-past association over a given time interval. The weak convergence of the underlying test statistic has been established, and several bootstrap approximations have been proposed.

The obtained simulation results suggest that the new test may be much more powerful than existing, less flexible methods, for special alternatives. This has been illustrated through real medical data analysis too. Specifically, we have analyzed the impact of chronic graft-versus-disease in evolution after a bone marrow

transplant in patients with leukemia, and also the relationship between the time to wound excision and time to Straphylocous Aureaus infection in burn patients. In both cases, the Markov condition for the corresponding illness-death model was tested, and the new method gave new interesting insights into the problem.

The relevance of the proposed test comes from the fact that no special dependence structure between the future and the past of the process is a priori assumed. Interestingly, a characterization of the Markov condition in terms of the markovianity of the censored, observable process has been given, in such a way that no Kaplan-Meier weights are needed for the computation of the proposed testing algorithms. Although a Kaplan-Meier-based formula for the association measure is possible, it is expected that it would lead to a less powerful test due to the large variance which would be typically obtained in heavily censored scenarios.

It would be interesting to extend the given methods to other more involved multi-state models. 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.

All the proposed methods were implemented in R. The code is available from the authors upon request.

Acknowledgements

Work supported by the Grants MTM2008-03129 of the Spanish Ministerio de Ciencia e Innovacion and 10PXIB300068PRof the Xunta de Galicia. Financial support from the INBIOMED project (DXPCTSUG, Ref. 2009/063) is also acknowledged.

References

[1] Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag, 1993.

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

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

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

[5] Frydman, H. Nonparametric estimation of a Markov illness-death process from interval-censored observations, with application to diabetes survival data. *Biometrika* 1995; **82**: 773–789.

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

[7] Harezlak, J., Gao, S. and Hui, S.L. An illness-death stochastic model in the analysis of longitudinal dementia data. *Statatistics in Medicine* 2003; **22**: 1465–1475.

[8] Frydman, H. and Szarek, M. Estimation of overall survival in an illness-death model with application to the vertical transmition of HIV-1. *Statatistics in Medicine* 2010; **29**: 2045–2054.

[9] Copelan, E.A., Biggs, J.C., Thompson, J.M., Crilley, P. et al. Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bu/Cy. *Blood* 1991; **78**: 838–843.

[10] Ichida, J.M., Wassell, J.T., Keller, M.D. and Ayers L.W. Evaluation of protocol change in burn-care management using the Cox proportional hazards model with time-dependent covariates. *Statistics in Medicine* 1993; **12**: 301–310.

[11] Meira-Machado, L., De Uña-Álvarez J. and Cadarso-Suárez, C. Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis* 2006; **12**: 325–344.

[12] Lin, D.Y., Sun, W. and Ying, Z. Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* 1999; **86**: 59–70.

[13] Wang, W. and Wells, M.T. Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* 1998; **85**: 561–572.

[14] Rodríguez-Girondo and de Uña-Álvarez. Testing markovianity in the three-state progressive model via future-past association. *Discussion Papers in Statistics and Operations Research 10/01, University of Vigo*;2010.

[15] Höffding, W. On the distribution of the rank correlation coefficient τ when the variates are not independent. *Biometrika* 1947; **34**: 183–196.

[16] Kruskal, W.H. Ordinal measures of association. *Journal of the American Statistical Association* 1958; **53**: 814–861.

[17] Serfling, R. J. *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.

[18] Billingsley, P. *Convergence of Probability Measures*. New York: Wiley, 1968.

Appendix: Technical proofs

Proof to Theorem 1. We begin stating two useful Lemmas. Let $b_H = \inf\{t : H(t) = 1\}$, where H is the df of $\widetilde{T}$. 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

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

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.$\blacksquare$

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.■

Proof to Theorem 2:

We want to prove that the process $\left\{\sqrt{n}(\widehat{\widetilde{\tau}}_t - \widetilde{\tau}_t) : t \in \mathcal{I}\right\}$ is tight, were $\mathcal{I}$ stands for an interval such that $P(\widetilde{A}_t) \geq c > 0$ for all $t \in \mathcal{I}$. Recall that

$$\widehat{\widetilde{\tau}}_t = \widehat{\widetilde{p}}_{c,t} - \widehat{\widetilde{p}}_{d,t}$$

where

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

and where $\widehat{\widetilde{F}}_t$ is the ordinary empirical df of the $\left(\widetilde{Z}_i, \widetilde{T}_i\right)$'s such that $\widetilde{Z}_i \leq t < \widetilde{T}_i$ and $\widehat{\widetilde{U}}_t(x,y) = \widehat{\widetilde{F}}_t(t, y^-) - \widehat{\widetilde{F}}_t(x^-, y^-)$. We first obtain an asymptotic representation of $\left\{\sqrt{n}(\widehat{\widetilde{\tau}}_t - \widetilde{\tau}_t) : t \in \mathcal{I}\right\}$ as a suitable process plus a remainder.

It is easily seen that

$$\widehat{\widetilde{p}}_{c,t} = \frac{2}{n_t(n_t - 1)} \sum_{i<j} I((\widetilde{Z}_i - \widetilde{Z}_j)(\widetilde{T}_i - \widetilde{T}_j) > 0) I(\widetilde{Z}_i \leq t < \widetilde{T}_i) I(\widetilde{Z}_j \leq t < \widetilde{T}_j)$$

where $n_t = \sum_{i=1}^n I(\widetilde{Z}_i \le t < \widetilde{T}_i)$. Since $n_t/n \to P(\widetilde{A}_t)$ as $n \to \infty$, we obtain

$$\widehat{\widetilde{p}}_{c,t} = \frac{2}{n(n-1)P(\widetilde{A}_t)^2} \sum_{i<j} I((\widetilde{Z}_i - \widetilde{Z}_j)(\widetilde{T}_i - \widetilde{T}_j) > 0) I(\widetilde{Z}_i \le t < \widetilde{T}_i) I(\widetilde{Z}_j \le t < \widetilde{T}_j) + o_P(1)$$

uniformly on $t \in \mathcal{I}$. A similar result holds for $\widehat{\widetilde{p}}_{d,t}$ and thus we have

$$\widehat{\widetilde{\tau}}_t = \frac{2}{n(n-1)P(\widetilde{A}_t)^2} \sum_{i \ne j} \varphi_t(\widetilde{Z}_i, \widetilde{Z}_j, \widetilde{T}_i, \widetilde{T}_j) - 1 + o_P(1)$$

where

$$\varphi_t(\widetilde{Z}_i, \widetilde{Z}_j, \widetilde{T}_i, \widetilde{T}_j) = I((\widetilde{Z}_i - \widetilde{Z}_j)(\widetilde{T}_i - \widetilde{T}_j) > 0) I(\widetilde{Z}_i \le t < \widetilde{T}_i) I(\widetilde{Z}_j \le t < \widetilde{T}_j).$$

Now, straightforward calculations show that the Hájek projection of the U-statistic

$$U_t = \frac{1}{n(n-1)} \sum_{i \ne j} \varphi_t(\widetilde{Z}_i, \widetilde{Z}_j, \widetilde{T}_i, \widetilde{T}_j)$$

is given by (cfr. [17], Section 5.3)

$$\widehat{U}_t = \frac{2P(\widetilde{A}_t)}{n} \sum_{i=1}^n \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_t(\widetilde{Z}_i, \widetilde{T}_i) \right] I(\widetilde{Z}_i \le t < \widetilde{T}_i) - P(\widetilde{A}_t)^2 \widetilde{p}_{c,t}$$

where $\widetilde{S}_t(x,y) = P(\widetilde{Z} > x, \widetilde{T} > y | \widetilde{A}_t)$. Since $U_t - \widehat{U}_t = o_P(n^{-1/2})$ we have

$$\begin{aligned}
\sqrt{n}(\widehat{\widetilde{\tau}}_t - \widetilde{\tau}_t) &= \sqrt{n} \left[\frac{2}{P(\widetilde{A}_t)^2} U_t - 1 - (2\widetilde{p}_{c,t} - 1) \right] + o_P(1) \\
&= \sqrt{n} \left[\frac{2}{P(\widetilde{A}_t)^2} (\widehat{U}_t - P(\widetilde{A}_t)^2 \widetilde{p}_{c,t}) \right] + o_P(1) \\
&= n^{-1/2} \frac{4}{P(\widetilde{A}_t)} \sum_{i=1}^n \left\{ \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_t(\widetilde{Z}_i, \widetilde{T}_i) \right] I(\widetilde{Z}_i \le t < \widetilde{T}_i) - P(\widetilde{A}_t) \widetilde{p}_{c,t} \right\} + o_P(1)
\end{aligned}$$

uniformly on $t \in \mathcal{I}$. Hence, tightness of $\sqrt{n}(\widehat{\widetilde{\tau}}_t - \widetilde{\tau}_t)$ follows provided that the process

$$R_n(t) = n^{-1/2} \sum_{i=1}^n \left\{ \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_t(\widetilde{Z}_i, \widetilde{T}_i) \right] I(\widetilde{Z}_i \le t < \widetilde{T}_i) - P(\widetilde{A}_t) \widetilde{p}_{c,t} \right\}$$

is tight. This is stated as a Lemma, which completes the proof.■

Lemma. The process $\{R_n(t) : t \in \mathcal{I}\}$ is tight.

Proof to Lemma. For $t_1 \leq t \leq t_2$ write

$$E\left\{|R_n(t) - R_n(t_1)|^2 |R_n(t_2) - R_n(t)|^2\right\} = E\left\{\left(n^{-1/2}\sum_{i=1}^{n}\alpha_i\right)^2\left(n^{-1/2}\sum_{i=1}^{n}\beta_i\right)^2\right\}$$

where

$$\alpha_i = \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_t(\widetilde{Z}_i, \widetilde{T}_i)\right] I(\widetilde{Z}_i \leq t < \widetilde{T}_i) - P(\widetilde{A}_t)\widetilde{p}_{c,t}$$
$$- \left[\widetilde{S}_{t_1}(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_{t_1}(\widetilde{Z}_i, \widetilde{T}_i)\right] I(\widetilde{Z}_i \leq t_1 < \widetilde{T}_i) + P(\widetilde{A}_{t_1})\widetilde{p}_{c,t_1}$$

and

$$\beta_i = \left[\widetilde{S}_{t_2}(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_{t_2}(\widetilde{Z}_i, \widetilde{T}_i)\right] I(\widetilde{Z}_i \leq t_2 < \widetilde{T}_i) - P(\widetilde{A}_{t_2})\widetilde{p}_{c,t_2}$$
$$- \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) + \widetilde{F}_t(\widetilde{Z}_i, \widetilde{T}_i)\right] I(\widetilde{Z}_i \leq t < \widetilde{T}_i) + P(\widetilde{A}_t)\widetilde{p}_{c,t},$$

$1 \leq i \leq n$. Note that $E(\alpha_i) = E(\beta_i) = 0$. Use a symmetry argument to write

$$E\left\{\left(n^{-1/2}\sum_{i=1}^{n}\alpha_i\right)^2\left(n^{-1/2}\sum_{i=1}^{n}\beta_i\right)^2\right\}$$

$$\begin{aligned} &= n^{-2}\left\{nE(\alpha_1^2\beta_1^2) + n(n-1)E(\alpha_1^2)E(\beta_1^2) + 2n(n-1)E^2(\alpha_1\beta_1)\right\} \\ &\leq n^{-2}\left\{nE(\alpha_1^2\beta_1^2) + 3n(n-1)E(\alpha_1^2)E(\beta_1^2)\right\}, \end{aligned}$$

the inequality following from Cauchy-Schwarz. Now, using $\widetilde{Z}_i \leq \widetilde{T}_i$ and $t_1 \leq t$ we get

$$\begin{aligned} I(\widetilde{Z}_i &\leq t_1 < \widetilde{T}_i) = I(\widetilde{T}_i > t_1) - I(\widetilde{Z}_i > t_1) \qquad (1) \\ &= I(\widetilde{T}_i > t) + I(t_1 < \widetilde{T}_i \leq t) - I(\widetilde{Z}_i > t) - I(t_1 < \widetilde{Z}_i \leq t) \\ &= I(\widetilde{Z}_i \leq t < \widetilde{T}_i) + I(t_1 < \widetilde{T}_i \leq t) - I(t_1 < \widetilde{Z}_i \leq t). \end{aligned}$$

Hence,

$$\alpha_i = \left[\widetilde{S}_t(\widetilde{Z}_i, \widetilde{T}_i) - \widetilde{S}_{t_1}(\widetilde{Z}_i, \widetilde{T}_i)\right] I(\widetilde{Z}_i \leq t < \widetilde{T}_i)$$

$$+\left[\widetilde{F}_t(\widetilde{Z}_i,\widetilde{T}_i)-\widetilde{F}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)\right]I(\widetilde{Z}_i\le t<\widetilde{T}_i)$$

$$-\left[\widetilde{S}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)+\widetilde{F}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)\right]I(t_1<\widetilde{T}_i\le t)+\left[\widetilde{S}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)+\widetilde{F}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)\right]I(t_1<\widetilde{Z}_i\le t)$$

$$-P(\widetilde{A}_t)\widetilde{p}_{c,t}+P(\widetilde{A}_{t_1})\widetilde{p}_{c,t_1},$$

while for the last term, using again (1), we obtain

$$-P(\widetilde{A}_t)\widetilde{p}_{c,t}+P(\widetilde{A}_{t_1})\widetilde{p}_{c,t_1}=P(\widetilde{A}_t)\left[\widetilde{p}_{c,t_1}-\widetilde{p}_{c,t}\right]+\widetilde{p}_{c,t_1}P(t_1<\widetilde{T}\le t)-\widetilde{p}_{c,t}I(t_1<\widetilde{Z}\le t).$$

Now, noting that $\widetilde{p}_{c,t}=\frac{1}{2}E\left[\widetilde{F}_t(\widetilde{Z},\widetilde{T})I(\widetilde{Z}\le t<\widetilde{T})\right]/P(\widetilde{A}_t)$, we have

$$\begin{aligned}2(\widetilde{p}_{c,t_1}-\widetilde{p}_{c,t}) &= \frac{1}{P(\widetilde{A}_{t_1})}E\left[\widetilde{F}_{t_1}(\widetilde{Z},\widetilde{T})(I(\widetilde{Z}\le t<\widetilde{T})+I(t_1<\widetilde{T}\le t)-I(t_1<\widetilde{Z}\le t))\right]\\ &\quad -\frac{1}{P(\widetilde{A}_{t_1})}E\left[\widetilde{F}_t(\widetilde{Z},\widetilde{T})I(\widetilde{Z}\le t<\widetilde{T})\right]\\ &\quad -(\frac{1}{P(\widetilde{A}_t)}-\frac{1}{P(\widetilde{A}_{t_1})})E\left[\widetilde{F}_t(\widetilde{Z},\widetilde{T})I(\widetilde{Z}\le t<\widetilde{T})\right]\end{aligned}$$

$$\begin{aligned}&= \frac{1}{P(\widetilde{A}_{t_1})}E\left[(\widetilde{F}_{t_1}(\widetilde{Z},\widetilde{T})-\widetilde{F}_t(\widetilde{Z},\widetilde{T}))I(\widetilde{Z}\le t<\widetilde{T})\right]\\ &\quad +\frac{1}{P(\widetilde{A}_{t_1})}E\left[\widetilde{F}_{t_1}(\widetilde{Z},\widetilde{T})(I(t_1<\widetilde{T}\le t)-I(t_1<\widetilde{Z}\le t))\right]\\ &\quad -\frac{P(\widetilde{A}_{t_1})-P(\widetilde{A}_t)}{P(\widetilde{A}_t)P(\widetilde{A}_{t_1})}E\left[\widetilde{F}_t(\widetilde{Z},\widetilde{T})I(\widetilde{Z}\le t<\widetilde{T})\right].\end{aligned}$$

Since (1) implies for some constant $k>0$

$$\left|\widetilde{S}_t(\widetilde{Z}_i,\widetilde{T}_i)-\widetilde{S}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)\right|\le k\left[P(t_1<\widetilde{T}\le t)+P(t_1<\widetilde{Z}\le t)\right]=k\left[\Gamma(t)-\Gamma(t_1)\right]$$

and

$$\left|\widetilde{F}_t(\widetilde{Z}_i,\widetilde{T}_i)-\widetilde{F}_{t_1}(\widetilde{Z}_i,\widetilde{T}_i)\right|\le k\left[\Gamma(t)-\Gamma(t_1)\right]$$

where $\Gamma(t)=P(\widetilde{T}\le t)+P(\widetilde{Z}\le t_1)$ is a continuous, nondecreasing function, and since $\widetilde{F}_t,\widetilde{S}_t\in[0,1]$, from $(a+b)^2\le 2(a^2+b^2)$ we have for some constant $k'>0$

$$E(\alpha_i^2)\le k'\left\{\left[\Gamma(t)-\Gamma(t_1)\right]^2+\Gamma(t)-\Gamma(t_1)\right\}.$$

An analogous result can be proved for β_i and hence

$$E(\alpha_1^2)E(\beta_1^2) \leq k'' \left\{ [\Gamma(t_2) - \Gamma(t_1)]^4 + [\Gamma(t_2) - \Gamma(t_1)]^2 \right\}.$$

Similarly, one may obtain $E(\alpha_1^2\beta_1^2) \leq k''' [\Gamma(t_2) - \Gamma(t_1)]^2$ and the result follows from Theorem 15.6 in [18].■