

# A Reproducing-Kernel-Hilbert-Space log-rank test for the two-sample problem

Tamara Fernández\*      Nicolás Rivera†

May 1, 2020

## Abstract

Weighted log-rank tests are arguably the most widely used tests by practitioners for the two-sample problem in the context of right-censored data. Many approaches have been considered to make weighted log-rank tests more robust against a broader family of alternatives, among them, considering linear combinations of weighted log-rank tests, and taking the maximum among a finite collection of them. In this paper, we propose as test statistic the supremum of a collection of (potentially infinite) weight-indexed log-rank tests where the index space is the unit ball in a reproducing kernel Hilbert space (RKHS). By using some desirable properties of RKHSs we provide an exact and simple evaluation of the test statistic and establish connections with previous tests in the literature. Additionally, we show that for a special family of RKHSs, the proposed test is omnibus. We finalise by performing an empirical evaluation of the proposed methodology and show an application to a real data scenario. Our theoretical results are proved using techniques for double integrals with respect to martingales that may be of independent interest.

**Key words:** Survival Analysis, Right-Censored Data, Reproducing Kernel Hilbert Space, Log-rank Test, two-sample tests.

## 1 Introduction

Two-sample testing is a classical problem in the context of survival data. For instance, in a clinical trial, two-sample tests can be used to compare different treatments when the survival times of patients are censored. Within the context of right-censored data, the classical log-rank test, first introduced by [Mantel \[1966\]](#) and [Peto and Peto \[1972\]](#), is the most widely used test among practitioners. A well-known property of the classical log-rank test is that it is the most powerful test under the assumption of proportional-hazard alternatives. This result can be deduced by noticing that the log-rank test statistic coincides with the score test statistic when the true cumulative hazard function belongs to the model  $\{\Lambda_\theta, \theta \in \Theta\}$ , where  $\Lambda_\theta(t) = \int_0^t e^\theta d\Lambda_0(x)$ ,  $\Lambda_0$  denotes the cumulative hazard function under the null hypothesis, and  $\theta \in \Theta \subseteq \mathbb{R}$  is the parameter of the model. While the classical log-rank test is optimal for proportional-hazard alternatives, it can have a substandard behaviour when the true cumulative hazard function cannot be expressed in terms of  $\{\Lambda_\theta, \theta \in \Theta\}$ .

In order to broaden the power of the classical log-rank test to other families of alternatives, researchers have introduced and studied different variants of weighted log-rank tests [[Tarone and Ware, 1977](#), [Gill, 1980](#), [Harrington and Fleming, 1982](#), [Bagdonavicius et al., 2010](#), [Andersen et al., 2012](#)]. We refer the reader to Chapter 7 of [Klein and Moeschberger \[2006\]](#) for a general discussion and comparison of different weighted log-rank tests approaches. In the simplest case, each weighted log-rank test with weight function  $\omega$  can be written as a score test for alternatives in the model  $\{\Lambda_{\theta,\omega}(t), \theta \in \Theta\}$ , where  $\Lambda_{\theta,\omega}(t) = \int_0^t e^{\theta\omega(F_0(x))} d\Lambda_0(x)$  and  $\omega$  is fixed. Then, similarly to the above, it can be deduced that each weighted log-rank test is the most powerful test for the null hypothesis  $H_0 : \Lambda_1 = \Lambda_0$  under the assumption that  $\Lambda_1$  can be expressed in terms of  $\{\Lambda_{\theta,\omega}, \theta \in \Theta\}$ , for some  $\theta \neq 0$ . If the true cumulative hazard function cannot be expressed in terms of this parametric model, there are no guarantees at all for the behaviour of the weighted log-rank test. Indeed, it may happen that we observe pathological cases in which the test has zero asymptotic power for specific alternatives. This is the case of the classical log-rank test, which is recovered by choosing  $\omega = 1$ , in the setting of crossing-hazard alternatives.

In an attempt to overcome the previously described limitations of weighted log-rank tests, researchers have considered two natural approaches to improve their performance: the first approach considered is to learn the weight function from the data, which defines a weighted log-rank test with an adaptive weight function, and the second approach consists on combining several log-rank tests into a single test statistic.

\*Gatsby Computational Neuroscience Unit, University College London. [t.a.fernandez@ucl.ac.uk](mailto:t.a.fernandez@ucl.ac.uk)

†Computer Laboratory, University of Cambridge. [nicolas.rivera@cl.cam.ac.uk](mailto:nicolas.rivera@cl.cam.ac.uk)

Both approaches avoid making the strong parametric assumption that a particular model is true. The first approach has been discussed by [Lai et al. \[1991\]](#), [Yang et al. \[2005\]](#) and [Yang and Prentice \[2010\]](#), and the second approach, which is the focus of this paper, has been studied by several authors. Particularly, [Tarone \[1981\]](#), [Kössler \[2010\]](#), and [Garès et al. \[2015\]](#) combined weighted log-rank tests by considering the maximum of a finite collection of them, and [Kosorok and Lin \[1999\]](#) proposed the supremum of an infinite collection of weight-indexed log-rank tests, with weights belonging to a particular space of functions. The latter approach is very close to what we propose in this paper, however, the test statistic proposed by [Kosorok and Lin \[1999\]](#) lacks of an analytically tractable expression, which forces the authors to rely on a Monte-Carlo approximation of it, whereas our approach, by an appropriate selection of the space of functions, obtains a simple expression for our test statistic, leading to a simple testing procedure.

While the focus of this paper are weighted log-rank test and the combination of them, there are other test-approaches for the two-sample problems that are worth mentioning. One example of these approaches are the weighted Kaplan-Meier estimators [[Pepe and Fleming, 1989](#)], which are parametrised on a weight function, and they can also be combined into one single statistic [[Shen and Cai, 2001](#)]. A different approach is the so-called empirical likelihood approach [[Zhou, 2016](#)], which leads to very interesting test statistics that can also be combined into a single statistic [[Bathke et al., 2009](#)]. Somewhat classical options are the Cramer von-Mises and the Kolmogorov-Smirnov test statistics, that have been deeply studied in many settings, including Survival Analysis [[Koziol and Green, 1976](#), [Koziol and Yuh, 1982](#)].

In this paper, we consider the supremum of a potentially infinite collection of weighted log-rank tests with weights belonging to the unit ball of a reproducing kernel Hilbert space of functions. Then, we propose to test the null hypothesis  $H_0 : \Lambda_1 = \Lambda_0$ , by checking if  $\sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}} \leq 1} \text{LR}_n(\omega) \approx 0$ , where  $\text{LR}_n(\omega)$  denotes a weighted log-rank test with weight function  $\omega$ , and  $\mathcal{H}$  is a reproducing kernel Hilbert space of functions. Clearly, if  $\Lambda_1$  can be expressed in terms of  $\{\Lambda_{\theta, \omega}, \theta \in \Theta, \omega \in B_1(\mathcal{H})\}$  for some  $\theta \neq 0$  and some  $\omega \in B_1(\mathcal{H})$  (where  $B_1(\mathcal{H})$  denotes the unit ball of  $\mathcal{H}$ ), our test statistic should be strictly greater than zero, in which case we decide against the null hypothesis. Also, notice that, the larger the space of functions  $\mathcal{H}$ , the more models we test at the same time.

Choosing weights  $\omega$  in the unit ball of a reproducing kernel Hilbert space  $\mathcal{H}$  is the key step in our testing procedure, as we will show that, by doing so, our test statistic  $\sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}} \leq 1} \text{LR}_n(\omega)$  can be evaluated exactly by a straightforward computation. This result follows from the good properties of reproducing kernel Hilbert spaces. Particularly, we will use two properties of reproducing kernel Hilbert spaces: i) they are uniquely characterised by a kernel function  $K$ , i.e., a symmetric and positive-definite function, and ii) they satisfy the so-called *reproducing kernel* property, which states that  $\omega(x) = \langle \omega, K(x, \cdot) \rangle_{\mathcal{H}}$  holds for any function  $\omega \in \mathcal{H}$  and any point  $x$ . By using this property we will show that  $\sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}} \leq 1} \text{LR}_n(\omega)^2 = \sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) c_i c_j$ , for some points  $x_i$  and constants  $c_j$  (depending on the data). This type of kernel estimators have been previously addressed by [Neuhaus \[1987\]](#) in connection with infinite sums of linear rank estimators (such as the log-rank) in the uncensored setting. Since, in practice, our test statistic is a quadratic form based on the kernel  $K$ , our method can be described just in terms of  $K$  avoiding the need to describe a Hilbert space. Describing kernel functions with interesting properties and useful interpretation has been a topic of extensive research, especially in the Machine Learning community. A compendium of popular kernels used in applications can be found in [Sousa \[2010\]](#).

While we can easily establish connections between our method and approaches based on weighted log-rank tests, we will show that an alternative interpretation of our test statistic allows us to connect our approach with the *kernel mean embedding* testing approach. In particular, we will show that our test statistic  $\sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}} \leq 1} \text{LR}_n(\omega)$  can be alternatively written as the norm of the difference of particular embeddings of our two samples into  $\mathcal{H}$ . The idea of comparing probability distributions/datasets by embedding them into a reproducing kernel Hilbert space of functions has been extensively studied in the uncensored setting by researchers in Statistics and Machine Learning [[Berlinet and Thomas-Agnan, 2004](#), [Smola et al., 2007](#)], however, it seems this idea has yet to percolate into the Survival Analysis community and, up to the best of our knowledge, it has only been considered by [Fernandez and Gretton \[2019\]](#) for the goodness-of-fit problem under right-censored data. In the uncensored case, two-sample tests were proposed and studied by [Gretton et al. \[2012\]](#). In this work, the authors embed two different samples into a reproducing kernel Hilbert space by considering the so-called “*kernel mean embedding*” of the empirical distribution of each sample. Then, they compute the difference between these two samples by computing the distance (induced by the norm) of the embeddings into  $\mathcal{H}$ . This idea can be straightforwardly applied to right-censored data by considering the kernel mean embedding of the Kaplan-Meier estimator instead of the empirical distribution, which leads to test statistics that are Kaplan-Meier  $V$ -statistics. These type of statistics were studied by [Bose and Sen \[2002\]](#) and by [Fernández and Rivera \[2020\]](#), and include as particular cases well-known statistics such as the Cramer von-Mises test statistic. As it was pointed out by [Fernández and Rivera \[2020\]](#) such an approach is not suitable for censored data as Kaplan-Meier  $V$ -statistics are only reliable when the amount of censored observations is rather small when compared to the total amount of data, and even in that case, the test statistic may not be data-efficient.

In this work, we study several asymptotic properties of our test statistic under both, the null and alternative hypothesis. Under the null hypothesis, we find the limiting distribution of our test statistic

by approximating it by a degenerate V-statistic. Under the alternative hypothesis, we prove that under reasonable conditions, our test has asymptotic power tending to one. While the asymptotic null distribution is known, in most cases, its parameters are hard to compute. Thus, in order to implement our test, we propose a Wild Bootstrap approximation of the null distribution and prove that the Wild Bootstrap statistic converges in distribution to our test statistic under the null hypothesis, giving us a correct Type-I error. Finally, we show that particular instances of our test statistic recover some existing testing approaches studied in the literature, which suggests our testing procedure is a natural generalisation of them. Examples of tests which our testing procedure recovers are: weighted log-rank tests, Pearson-type tests [Akritas, 1988] and projection-type tests [Brendel et al., 2014, Ditzhaus and Friedrich, 2018].

Besides theoretical guarantees, we provide an extensive empirical evaluation of our method in two important data scenarios: proportional hazard functions and time-dependent hazard functions, including Weibull and periodic hazard alternatives. In our experiments we demonstrate that our method has a good performance in a wide range of problems, which include sample sizes from small to large, and different censoring percentages. Our experiments show that finite-dimensional reproducing kernel Hilbert spaces tend to have an overall better performance in problems with fewer observations or a simpler hazard function. While more complex reproducing kernel Hilbert spaces (i.e., infinite-dimensional spaces) still show a good performance in these simpler problems, our experiments show that they are better suited for larger datasets and more complex hazard functions. We also provide a real-data evaluation of our method, trying different kernel functions, and we compare the results with those obtained by Ditzhaus and Friedrich [2018].

The structure of the paper is as follows. In Section 2, we introduce some standard notation used in Survival Analysis and describe weighted log-rank tests. In Section 3 we introduce the essential background knowledge about reproducing kernel Hilbert spaces (RKHS) and formally introduce our test statistic. In Section 4 we study asymptotic properties of our test. Later, in Section 5, we proceed to explain how to implement a Wild Bootstrap approach. Sections 6 and 7 are devoted to empirical studies using simulated data and real data, respectively.

## 2 Survival analysis background

### 2.1 General notation

We establish some general notation that will be used throughout the paper. We denote  $\mathbb{R}_+ = (0, \infty)$ . Let  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be an arbitrary right-continuous function, then we define  $f(x-) = \lim_{h \rightarrow 0} f(x - |h|)$ . In this work we make use of standard asymptotic notation [Janson, 2011], e.g.,  $O_p$ ,  $o_p$ ,  $\Theta_p$ , etc., with respect to the number of sample points  $n$ . In order to avoid large parentheses, we write  $X = O_p(1)Y$  instead of  $X = O_p(Y)$ , especially if the expression for  $Y$  is very long. Given a sequence of stochastic processes  $(W_n(x) : x \in \mathcal{X})$ , depending on the number of observations  $n$ , and a function  $f(x) \geq 0$ , we say that  $W_n(x) = O_p(1)f(x)$  uniformly for all  $x \in A_n$ , if and only if  $\sup_{x \in A_n} |W_n(x)|/f(x) = O_p(1)$ , where  $A_n \subseteq \mathcal{X}$  is a set that may depend on  $n$ , and we use the convention  $0/0 = 0$ . Lastly, in this work the integral symbol  $\int_a^b$  means integration over  $(a, b]$  unless we state otherwise.

### 2.2 Right-censored data

Our data corresponds to right-censored observations belonging to two groups/classes, namely group 0 and group 1. We assume that the total number of observations is  $n$ , and that  $n = n_0 + n_1$ , where  $n_c$  denotes the number of observations in group  $c \in \{0, 1\}$ . We use  $[n]$  to denote the set  $\{1, \dots, n\}$ . For the sake of asymptotic analysis, we assume that we have no vanishing groups, that is,  $n_0/n \rightarrow \eta$  and  $n_1/n \rightarrow (1 - \eta)$  with  $\eta \in (0, 1)$  as  $n$  tends to infinity.

Right-censored datasets are commonly observed in triples  $((X_i, \Delta_i, c_i))_{i \in [n]}$ , where  $X_i = \min\{T_i, C_i\}$  is an *observed time*, defined as the minimum between a *survival time* of interest  $T_i$  and a *censoring time*  $C_i$ , the variable  $\Delta_i = \mathbb{1}_{\{X_i = T_i\}}$  is an indicator of whether we actually observe the survival time of interest, and  $c_i \in \{0, 1\}$  is an associated covariate that denotes the group membership of the  $i$ -th observation. In this work we assume that all triples are mutually independent and that the triples have the same distribution within each group. Additionally, we assume *non-informative censoring*, meaning that, given the covariate  $c_i$ , the survival time  $T_i$  is independent of the censoring time  $C_i$ .

We denote by  $F_c$ ,  $G_c$ , and  $H_c$ , the respective conditional distribution functions of the random variables  $T_i$ ,  $C_i$ , and  $X_i$ , given that the covariate  $c_i$  takes the value  $c \in \{0, 1\}$ . Notice that  $1 - H_c = (1 - F_c)(1 - G_c)$  due to the non-informative censoring assumption. For simplicity of exposition, we assume that the distribution functions  $F_c$  and  $G_c$  are continuous distributions on  $\mathbb{R}_+$ , however, all the methods of this paper can be extended to general probability distributions on  $\mathbb{R}$ , as our results are based on counting processes arguments that have been developed in full generality. We denote by  $S_c(x) = 1 - F_c(x)$  and by  $\Lambda_c(x) = \int_{(0, x]} S_c(s)^{-1} dF_c(s)$ , the so-called *survival* and *cumulative hazard* functions associated with  $F_c$ , respectively, and we denote by  $\widehat{S}_c$  the Kaplan-Meier estimator of  $S_c$ , and by  $\widehat{\Lambda}_c$  the Nelson-Aalen estimator of  $\Lambda_c$ . Notice that for the computation of  $\widehat{S}_c$  and  $\widehat{\Lambda}_c$  we only use triples  $(X_i, \Delta_i, c_i)$  satisfying  $c_i = c$ .

While in principle we assume that the covariates  $c_i$  are deterministic, it is also possible to consider them as independent and identically distributed (i.i.d.) samples from a Bernoulli distribution with mean  $\eta$  (notice that in this case  $n_0$  is a random variable satisfying  $n_0/n \rightarrow \eta$  almost surely). These two possible assumptions originate two models: the *deterministic covariates* model and the *random covariates* model. For our analysis, it will be convenient to use the random covariates model as, under this assumption, the triples  $((X_i, \Delta_i, c_i))_{i \in [n]}$  are i.i.d. which simplifies the asymptotic analysis, nevertheless, as we will show in Appendix B.3, both models are asymptotically equivalent. Also, in practice, there is no difference between the two models when implementing our testing procedure.

In this paper, we use the notion of pooled data and pooled distributions. The pooled data corresponds to the original data after ignoring/removing the covariates  $(c_i)_{i \in [n]}$ , that is, the pooled data is  $((X_i, \Delta_i))_{i \in [n]}$ . We use the term pooled distributions to refer to the distributions of a randomly selected data-point. Particularly, under the random covariates model, the pooled distributions associated with the observed times  $(X_i)_{i \in [n]}$  and the survival times  $(T_i)_{i \in [n]}$ , are the marginal distributions of  $X_i$  and  $T_i$ , respectively, given by  $H = \eta H_0 + (1 - \eta)H_1$  and  $F = \eta F_0 + (1 - \eta)F_1$ . Under the deterministic covariates model, the pooled distributions associated with the observed times and the survival times correspond to  $(n_0/n)H_0 + (n_1/n)H_1$  and  $(n_0/n)F_0 + (n_1/n)F_1$ , respectively. Notice that, since  $(n_0/n)H_0 + (n_1/n)H_1 \rightarrow H$  and  $(n_0/n)F_0 + (n_1/n)F_1 \rightarrow F$  as  $n$  grows to infinity, the pooled distributions of both models are asymptotically equivalent. Additionally, we denote by  $S(x) = 1 - F(x)$  and by  $\Lambda(x) = \int_{(0,x]} S(s)^{-1} dF(s)$  the survival and cumulative hazard functions associated with the pooled distribution  $F$ , and we use the Kaplan-Meier estimator  $\widehat{S}$  and the Nelson-Aalen estimator  $\widehat{\Lambda}$  to approximate  $S$  and  $\Lambda$ , respectively. Notice that  $\widehat{S}$  and  $\widehat{\Lambda}$  are computed using the pooled-data.

## 2.3 Counting processes

In this work, we use the standard counting processes notation used in Survival Analysis. We define the *individual, class and pooled* counting processes by  $N^i(x) = \Delta_i \mathbb{1}_{\{X_i \leq x\}}$  for  $i \in [n]$ ,  $N_c(x) = \sum_{i=1}^n \mathbb{1}_{\{c_i=c\}} N^i(x)$  for  $c \in \{0, 1\}$ , and  $N(x) = \sum_{i=1}^n N^i(x) = N_0(x) + N_1(x)$ , respectively. Similarly, we define the individual, class and pooled risk functions by  $Y^i(x) = \mathbb{1}_{\{X_i \geq x\}}$ ,  $Y_c(x) = \sum_{i=1}^n \mathbb{1}_{\{c_i=c\}} Y^i(x)$  and  $Y(x) = \sum_{i=1}^n Y^i(x) = Y_0(x) + Y_1(x)$ , respectively. By using the previous notation, we write the Nelson-Aalen estimator of  $d\Lambda_c$  as  $d\widehat{\Lambda}_c(x) = dN_c(x)/Y_c(x)$ , and the pooled Nelson-Aalen estimator as  $d\widehat{\Lambda}(x) = dN(x)/Y(x)$ . To avoid ambiguities we assume that  $0/0 = 0$ , e.g.,  $d\widehat{\Lambda}(x) = 0$  if  $Y(x) = 0$ . Finally, we define the process  $L(x) = Y_0(x)Y_1(x)/Y(x)$ , which appears frequently in our results, and we define  $\tau_n = \max\{X_i : i \in [n]\}$  and  $\tau = \sup\{x : 1 - H(x) > 0\}$ . Observe that  $\tau_n \rightarrow \tau$  a.s. and that  $N^i(x) = N^i(\tau_n)$ ,  $N_c(x) = N_c(\tau_n)$  and  $N(x) = N(\tau_n)$  for any  $x \geq \tau_n$ . Also, notice that all our observed times,  $X_i$ , are less than  $\tau$  as they are generated from continuous distributions.

We assume that all random variables are defined on a common filtrated probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_x)_{x \geq 0}, \mathbb{P})$ , where the sigma-algebra  $\mathcal{F}_x$  is generated by

$$\{\mathbb{1}_{\{X_i \leq s, \Delta_i=0\}}, \mathbb{1}_{\{X_i \leq s, \Delta_i=1\}} : 0 \leq s \leq x, i \in [n]\},$$

and the  $\mathbb{P}$ -null sets of  $\mathcal{F}$ . Under the deterministic covariates model, we define the individual  $(\mathcal{F}_x)$ -martingales associated to each data point,  $(X_i, \Delta_i, c_i)$ , as  $M^i(x) = N^i(x) - \int_{(0,x]} Y^i(s) d\Lambda_{c_i}(s)$ . Similarly, under the deterministic covariates model, we define the class and pooled martingales by  $M_c(x) = \sum_{i=1}^n \mathbb{1}_{\{c_i=c\}} M^i(x)$  and  $M(x) = \sum_{i=1}^n M^i(x) = M_0(x) + M_1(x)$ , respectively, where  $c \in \{0, 1\}$ . Notice that under the random covariates model,  $M^i(x)$ ,  $M_c(x)$  and  $M(x)$  are  $(\mathcal{F}_x)$ -martingales when conditioning on  $(c_i)_{i \in [n]}$  (alternatively, we can include the sigma algebra associated with  $(c_i)_{i \in [n]}$  to generate  $\mathcal{F}_x$ ). We denote by  $\langle M^i \rangle(x)$  and  $[M^i](x)$ , respectively, the predictable and quadratic variation processes associated with  $M^i(x)$ , which, in the context of continuous survival and censoring times, are given by  $d\langle M^i \rangle(x) = Y^i(x) d\Lambda_{c_i}(x)$  and  $d[M^i](x) = dN^i(x)$ . For further information about counting processes and their applications in Survival Analysis we refer the reader to Fleming and Harrington [1991].

Notice that, since we consider continuous survival times, we can estimate  $F(x)$  by either  $\widehat{F}(x)$  or  $\widehat{F}(x-)$ . The main advantage of using the latter estimator is that  $(\widehat{F}(x-))_{x \geq 0}$  is left-continuous, and thus, it is a predictable process with respect to the filtration  $(\mathcal{F}_x)_{x \geq 0}$ .

## 2.4 The log-rank estimator

The weighted log-rank statistic is defined by

$$\text{LR}_n(\omega) = \frac{n}{n_0 n_1} \int_0^{\tau_n} \omega(\widehat{F}(x-)) L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right), \quad (1)$$

where the function  $\omega : [0, 1] \rightarrow \mathbb{R}$  is referred to as *weight function*. In the previous equation recall that  $L(x) = Y_0(x)Y_1(x)/Y(x)$ ,  $\widehat{F}$  is the pooled Kaplan-Meier estimator of the pooled distribution function  $F$ , and  $\widehat{\Lambda}_0$  and  $\widehat{\Lambda}_1$  are the class Nelson-Aalen estimators of the cumulative hazard functions  $\Lambda_0$  and  $\Lambda_1$ , respectively.

Gill [1980] studied the asymptotic behaviour of the weighted log-rank statistic,  $\text{LR}_n(\omega)$ , under the null hypothesis  $H_0 : F_0 = F_1$ , obtaining that

$$\sqrt{\frac{n_0 n_1}{n}} \text{LR}_n(\omega) \xrightarrow{\mathcal{D}} N(0, \sigma^2), \quad (2)$$

where

$$\sigma^2 = \int_0^\tau w(F_0(x))^2 \frac{(1 - G_0(x))(1 - G_1(x))}{\eta(1 - G_0(x)) + (1 - \eta)(1 - G_1(x))} dF_0(x).$$

Notice that, even under the null hypothesis, the censoring distributions  $G_0$  and  $G_1$  are not necessarily equal, hence the expression given for  $\sigma^2$  cannot be simplified. Also, in most cases, the censoring distributions are unknown, and thus it is not possible to evaluate  $\sigma^2$  to find rejection regions. To carry-out a testing procedure, we can use the asymptotic null distribution given in Equation (2), replacing the asymptotic variance  $\sigma^2$  by  $\hat{\sigma}^2 = n/(n_0 n_1) \int_0^{\tau_n} \omega(\hat{F}(x-))^2 L(x) d\hat{\Lambda}(x)$ , which is an unbiased estimator of  $\sigma^2$  under the null hypothesis. We refer the reader to Gill [1980] for more details.

It is a well-known fact that weighted log-rank tests relate to score tests through the choice of a specific family of alternatives. Let

$$\Lambda(x; \theta, \omega) = \int_0^x e^{\theta \omega(F_0(s))} d\Lambda_0(s), \quad \theta \in \Theta, \quad (3)$$

be a parametric model (indexed by  $\theta$ ) for the cumulative hazard function, where  $\omega : [0, 1) \rightarrow \mathbb{R}$  is a fixed continuous function, and  $\Theta$  is chosen as an open subset of  $\mathbb{R}$  containing 0. Then, under the assumption that  $\Lambda_1(x) = \Lambda(x; \theta_1, \omega)$  for some  $\theta_1 \in \Theta$ , testing the null hypothesis,  $H_0 : \Lambda_0 = \Lambda_1$ , is equivalent to test  $H_0 : \theta = 0$ .

We can test  $H_0 : \theta = 0$  by using a score test. The score statistic is computed in terms of the score function defined by  $U(\theta) = \partial/\partial\theta \log L_n(\theta; \omega)$ , where  $L_n(\theta; \omega)$  is the likelihood function under the model of Equation (3). For the Goodness-of-Fit problem, where  $F_0$  and  $\Lambda_0$  are known, the score statistic is given by

$$U(0) = \int_0^{\tau_n} w(F_0(x)) Y_1(x) \left( d\Lambda_0(x) - d\hat{\Lambda}_1(x) \right).$$

In the Two-Sample problem,  $\Lambda_0$  and  $F_0$  are unknown, but they can be estimated using the **pooled** Nelson-Aalen and Kaplan-Meier estimators, obtaining

$$\hat{U}(0) = \int_0^{\tau_n} w(\hat{F}(x-)) L(x) \left( d\hat{\Lambda}(x) - d\hat{\Lambda}_1(x) \right).$$

Then, a simple comparison shows that  $n/(n_0 n_1) \hat{U}(0) = \text{LR}_n(\omega)$ , from which we deduce the relationship between the weighted log-rank test defined in Equation (1) and the score test associated to the parametric model defined in Equation (3). By the Neyman-Pearson Lemma, we deduce that the weighted log-rank test is the most powerful test for small departures from the null, that is, when  $\theta \rightarrow 0$ . Unfortunately, if the model in Equation (3) is misspecified, little can be said about the performance of the test, indeed, it is well-known that, in some cases, the weighted log-rank statistic yields a test with asymptotically zero-power.

### 3 An RKHS approach to the log-rank test

A standard approach to broaden the power of weighted log-rank tests, in order to achieve a more robust behaviour across a larger class of alternatives, is to combine several weighted log-rank tests into a single statistic. In this way, if one of the weight functions completely fails to differentiate between the null and alternative hypotheses, we can still rely on the remaining weight functions to help us to discriminate, and thus increase the overall power of our testing procedure. Two interesting questions that arise from this approach are: How do we combine these weighted log-rank tests? Which and how many weight functions do we need to choose?

Selecting and combining weight functions efficiently for the problem at hand is very difficult as, in most cases, it requires that the user analyses the data in advance to select appropriate weight functions, e.g., it is usual to check if the hazard functions cross or/and if they show late/early departures. Searching for important features in the data may be very time-consuming and, moreover, it is always possible that there are features that do not translate into an appropriate weight function. Also, it might happen that the user overlooks some important relations.

Instead of relying on the expertise of a user to identify important features, we can consider as many weight functions as possible to account for the heterogeneity in the data. We can take this approach to the extreme by considering a potentially infinite family of weights. While this approach solves the problem of choosing weight functions, naively choosing a particular family of weights will lead to tests that, i) are hard to

calibrate as we will need several data points to reach the correct Type-I error, ii) do not have an analytically tractable closed form for the test statistic, leading to tests that iii) are computationally expensive.

In order to overcome the previous difficulties, we consider weight functions belonging to the unit ball of a reproducing kernel Hilbert space  $\mathcal{H}$ , and propose as test statistic  $Z_n$ , which is defined by

$$Z_n = \sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}}^2 \leq 1} \text{LR}_n(\omega)^2. \quad (4)$$

While a priori there is not a good reason to choose this particular family of weights, we will show that it has very nice features that translate into desirable properties of our testing procedure; among others, the test statistic has a simple closed form, allowing simple computations as well as an economic Wild Bootstrap implementation.

### 3.1 Reproducing kernel Hilbert spaces

A reproducing kernel Hilbert space (RKHS) is a Hilbert space of functions  $\omega : [0, 1) \rightarrow \mathbb{R}$  satisfying that the evaluation  $\omega \rightarrow \omega(x)$  is continuous for every fixed  $x \in [0, 1)$  (it is worth mentioning that we can replace  $[0, 1)$  by any space). By the Riesz representation theorem, for any  $x \in [0, 1)$ , there exists a unique element  $K_x \in \mathcal{H}$  such that  $\omega(x) = \langle K_x, \omega \rangle_{\mathcal{H}}$  for all  $\omega \in \mathcal{H}$ , which is known as the *reproducing kernel property*. Since  $K_x \in \mathcal{H}$  for all  $x \in [0, 1)$ ,  $K_x(y) = K_y(x) = \langle K_x, K_y \rangle_{\mathcal{H}}$  holds for any  $x, y \in [0, 1)$ . Then, this result allows us to define the so-called *reproducing kernel*  $K : [0, 1)^2 \rightarrow \mathbb{R}$  as

$$K(x, y) = \langle K_x, K_y \rangle_{\mathcal{H}}. \quad (5)$$

From now on, in order to ease the notation, we write  $K(x, \cdot)$  instead of  $K_x(\cdot)$ , even though the former induces a slight abuse of notation.

For every RKHS  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  there exists a unique symmetric positive-definite reproducing kernel  $K : [0, 1)^2 \rightarrow \mathbb{R}$  satisfying Equation (5). Conversely, by the Moore-Aronszajn Theorem [Aronszajn, 1950], for any symmetric positive-definite kernel function  $K : [0, 1)^2 \rightarrow \mathbb{R}$ , there exists a unique RKHS  $\mathcal{H}$  for which  $K$  is its reproducing kernel.

Given a finite signed measure  $\mu$  on  $[0, 1)$ , we define the kernel mean embedding of  $\mu$  into  $\mathcal{H}$  as

$$\mu \rightarrow \int_{[0,1)} K(y, \cdot) \mu(dy) \in \mathcal{H}, \quad (6)$$

where the previous integral has to be understood as a Bochner integral. A sufficient condition to guarantee the existence of such an embedding is that  $\int_{[0,1)} \sqrt{K(x, x)} |\mu|(dx) < \infty$  [Gretton et al., 2012, Lemma 3]. A kernel  $K$  is said to be *characteristic* if the mean kernel embedding, defined in Equation (6), is injective on the space of probability distributions (i.e., distinct probability measures are embedded as different elements of  $\mathcal{H}$ ). Furthermore, a continuous kernel  $K$  is said to be *c-universal* if the mean kernel embedding is injective on the set of finite signed measures. Clearly, a continuous *c-universal* kernel is characteristic. Most of the standard kernels used in applications are *c-universal*, e.g., the Gaussian kernel,  $K(x, y) = \exp\{-(x-y)^2/\sigma^2\}$ , and the OrnsteinUhlenbeck kernel,  $K(x, y) = \exp\{-|x-y|/|\sigma|\}$ . For more information about characteristic and *c-universal* kernels we refer the reader to Fukumizu et al. [2009], Muandet et al. [2017], Simon-Gabriel and Schölkopf [2018] and references therein.

### 3.2 An RKHS log-rank test

Recall that

$$Z_n = \sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}}^2 \leq 1} \text{LR}_n(\omega)^2,$$

where  $\mathcal{H}$  is an RKHS of functions  $\omega : [0, 1) \rightarrow \mathbb{R}$  with reproducing kernel  $K : [0, 1)^2 \rightarrow \mathbb{R}$ . In this section, we show that it is possible to find a closed-form expression for our test statistic,  $Z_n$ , in terms of the reproducing kernel  $K$  associated with  $\mathcal{H}$ . This result is formally stated in the following theorem:

**Theorem 1.**

$$Z_n = \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} K(\hat{F}(x-), \hat{F}(y-)) L(x) L(y) \left( d\hat{\Lambda}_0(x) - d\hat{\Lambda}_1(x) \right) \left( d\hat{\Lambda}_0(y) - d\hat{\Lambda}_1(y) \right). \quad (7)$$

By using the definition of the class Nelson-Aalen estimators, Equation (7) can be rewritten as

$$Z_n = \left( \frac{n}{n_0 n_1} \right)^2 \sum_{i=1}^n \sum_{j=1}^n K(\hat{F}(X_i-), \hat{F}(X_j-)) L(X_i) L(X_j) (-1)^{c_i + c_j} \frac{\Delta_i}{Y_{c_i}(X_i)} \frac{\Delta_j}{Y_{c_j}(X_j)}, \quad (8)$$

which is a quadratic form with matrix-representation given by

$$Z_n = \left( \frac{n}{n_0 n_1} \right)^2 \mathbf{V}^\top \widehat{\mathbf{K}} \mathbf{V}, \quad (9)$$

where  $\mathbf{V} \in \mathbb{R}^n$  is defined by  $\mathbf{V}_j = L(X_j)(-1)^{c_j} \Delta_j / Y_{c_j}(X_j)$  and  $\widehat{\mathbf{K}} \in \mathbb{R}^n \times \mathbb{R}^n$  is the matrix defined by  $\widehat{\mathbf{K}}_{i,j} = K(\widehat{F}(X_i), \widehat{F}(X_j))$ .

In order to prove Theorem 1, we introduce an alternative embedding of the data, which is different to the kernel mean embedding of Equation (6). Given a measure  $\nu$  (not necessarily a probability measure), we define the embedding  $\phi$  of a finite signed measure  $\nu$  into  $\mathcal{H}$  by

$$\phi_\nu(\cdot) = \int_0^\tau K(F(y), \cdot) \nu(dy).$$

In practice, since the pooled distribution  $F$  is unknown, we replace it by the pooled Kaplan-Meier estimator  $\widehat{F}(x-)$  (recall that the survival time are continuous). Then, by using the previous definition, we introduce

$$\phi_0^n(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(y-), \cdot) L(y) d\widehat{\Lambda}_0(y), \quad \text{and} \quad \phi_1^n(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(y-), \cdot) L(y) d\widehat{\Lambda}_1(y), \quad (10)$$

which are the corresponding embeddings of two empirical measures,  $\nu_0^n$  and  $\nu_1^n$ , into  $\mathcal{H}$ , where  $\nu_0^n$  and  $\nu_1^n$  are defined for any  $s < t$  by

$$\nu_0^n((s, t]) = \frac{n}{n_0 n_1} \int_s^t L(x) d\widehat{\Lambda}_0(x), \quad \text{and} \quad \nu_1^n((s, t]) = \frac{n}{n_0 n_1} \int_s^t L(x) d\widehat{\Lambda}_1(x), \quad (11)$$

respectively. Notice that  $\phi_0^n$  and  $\phi_1^n$  are always well-defined, meaning that  $\phi_0^n \in \mathcal{H}$  and  $\phi_1^n \in \mathcal{H}$ , since they are just finite sums of elements of  $\mathcal{H}$  (observe that  $K(\widehat{F}(y-), \cdot) \in \mathcal{H}$  for any fixed  $y$ ).

By using the previous embeddings, we obtain an inner product representation of the weighted log-rank statistic, which will be used in the proof of Theorem 1.

**Lemma 2** (Log-rank representation). *For any  $\omega \in \mathcal{H}$ ,*

$$\text{LR}_n(\omega) = \langle \omega, \phi_0^n - \phi_1^n \rangle_{\mathcal{H}},$$

and then

$$Z_n = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2.$$

*Proof.* Since  $\widehat{F}(x-) \in [0, 1]$  for any  $x \leq \tau_n$ , the reproducing property yields  $\omega(\widehat{F}(x-)) = \langle \omega(\cdot), K(\widehat{F}(x-), \cdot) \rangle_{\mathcal{H}}$ . Then, by the linearity of the inner product and integration,

$$\begin{aligned} \text{LR}_n(\omega) &= \frac{n}{n_0 n_1} \int_0^{\tau_n} \omega(\widehat{F}(x-)) L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \\ &= \frac{n}{n_0 n_1} \int_0^{\tau_n} \langle \omega, K(\widehat{F}(x-), \cdot) \rangle_{\mathcal{H}} L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \\ &= \left\langle \omega, \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \right\rangle_{\mathcal{H}} \\ &= \langle \omega, \phi_0^n - \phi_1^n \rangle_{\mathcal{H}}. \end{aligned}$$

Finally, by taking supremum over the unit ball, we have that

$$Z_n = \sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}}^2 \leq 1} \text{LR}_n(\omega)^2 = \sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}}^2 \leq 1} \langle \omega, \phi_0^n - \phi_1^n \rangle_{\mathcal{H}}^2 = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2, \quad (12)$$

where the last equality holds since  $\mathcal{H}$  is a Hilbert space.  $\square$

*Proof of Theorem 1.* By Equation (10), we have  $(\phi_0^n - \phi_1^n)(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) (d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x))$ . Then, by using the linearity of the inner product and integration, and the reproducing kernel property, we deduce

$$\begin{aligned} &\|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 \\ &= \left\langle \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right), \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(y-), \cdot) L(y) \left( d\widehat{\Lambda}_0(y) - d\widehat{\Lambda}_1(y) \right) \right\rangle_{\mathcal{H}} \\ &= \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} \left\langle K(\widehat{F}(x-), \cdot), K(\widehat{F}(y-), \cdot) \right\rangle_{\mathcal{H}} L(x) L(y) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \left( d\widehat{\Lambda}_0(y) - d\widehat{\Lambda}_1(y) \right) \\ &= \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(y-)) L(x) L(y) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \left( d\widehat{\Lambda}_0(y) - d\widehat{\Lambda}_1(y) \right). \end{aligned}$$

$\square$

### 3.3 Recovering existing tests

Our testing approach is based on fixing an RKHS of functions  $\mathcal{H}$ , which is done by fixing a kernel  $K$ . We show that for specific choices of the reproducing kernel  $K$ , we recover some previously known tests.

#### 3.3.1 Weighted log-rank tests

In order to recover the classical weighted log-rank test  $\text{LR}_n(\omega)$ , we set  $K(x, y) = \omega(x)\omega(y)$ . Notice that  $K$  is symmetric and non-negative definite for any  $\omega$ . Then, by evaluating  $Z_n$  using Equation (7), we obtain

$$Z_n = \text{LR}_n(\omega)^2.$$

#### 3.3.2 Pearson-type tests

Consider the partition of the interval  $(0, 1]$  given by  $I_j = (j/k, (j+1)/k]$ , where  $j \in \{0, \dots, k-1\}$  and  $k > 0$  is an integer. We recover Pearson-type testing approaches by choosing the kernel  $K(x, y) = \sum_{j=0}^{k-1} \omega(x)\omega(y)\mathbb{1}_{I_j \times I_j}(x, y)$ , where  $\omega : [0, 1) \rightarrow \mathbb{R}$  is a given weight function. Then, by evaluating  $Z_n$  using Equation (7), we get

$$Z_n = \sum_{j=0}^{k-1} \text{LR}_n(\omega\mathbb{1}_{\{I_j\}})^2 = \sum_{j=0}^{k-1} \left( \int_{I_j} \omega(\widehat{F}(x-))L(x) \left( d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x) \right) \right)^2.$$

Particularly, if we choose  $\omega \equiv 1$ , we recover

$$Z_n = \sum_{j=0}^{k-1} \left( \sum_{X_i \in I_j} \frac{Y_1(X_i)}{Y(X_i)} \Delta_i \mathbb{1}_{\{c_i=0\}} - \sum_{X_i \in I_j} \frac{Y_0(X_i)}{Y(X_i)} \Delta_i \mathbb{1}_{\{c_i=1\}} \right)^2, \quad (13)$$

which compares observed frequencies on  $I_j$  between the two groups. We can go further and consider the normalised version of Pearson-type tests by choosing the kernel as

$$K(x, y) = \sum_{j=0}^{k-1} \frac{\omega(x)\omega(y)}{\sigma_j(\omega)^2} \mathbb{1}_{I_j \times I_j}(x, y),$$

where  $\sigma_j(\omega)^2 = \frac{n}{n_0 n_1} \int_{I_j} \omega(\widehat{F}(x-))^2 L(x) d\widehat{\Lambda}(x)$ . Notice that in this case  $K(x, y)$  is a random kernel, but it converges to a deterministic one since  $\sigma_j(\omega)$  converges almost surely to a constant when  $n$  tends to infinity.

#### 3.3.3 Projection-type test

Projection-type tests were introduced by [Brendel et al. \[2014\]](#), and recently revisited by [Ditzhaus and Friedrich \[2018\]](#). Consider a finite number of weight functions  $w_1, \dots, w_k$  such that  $w_i \circ F \in L^2(\nu)$  for all  $i \in \{1, \dots, k\}$ , where  $\nu$  is the measure given by

$$\nu(dx) = \frac{(1 - H_0(x))(1 - H_1(x))}{1 - H(x)} d\Lambda(x).$$

The projection-type testing approach considers the following test statistic

$$\sum_{i=1}^k \text{LR}(\tilde{w}_i)^2,$$

where  $\{\tilde{w}_i\}_{i=1}^k$  is an orthonormal base of the subspace  $U$  of  $L^2(\nu)$  generated by  $w_1 \circ F, \dots, w_k \circ F$ . We refer the reader to [Brendel et al. \[2014\]](#) for a detailed explanation. [Brendel et al. \[2014\]](#) and [Ditzhaus and Friedrich \[2018\]](#) recommend using weights with some meaning, for instance,  $\omega(x) = x - 1/2$  is used to detect a cross between two hazard functions around the median of the pooled survival time distribution, and  $\omega(x) = x^p(1-x)^q$  is used to detect early and/or late differences between the hazard functions, depending on the parameters  $p$  and  $q$ . Nevertheless, observe that, in terms of projections, the meaning of the functions does not matter as, for instance, projecting over the subspace generated by  $\{1, x - 1/2, x(1-x), x^2(1-x)\}$  is the same as projecting over the subspace generated by  $\{1, x, x^2, x^3\}$ .

Our approach can be seen as a natural generalisation of the previous method. Indeed, we can recover the previous test statistic by choosing the kernel

$$K(x, y) = \sum_{i=1}^k \tilde{w}_i(x)\tilde{w}_i(y).$$



To compute the orthonormal basis  $\{\tilde{\omega}_i\}_{i=1}^k$ , we just need to compute the  $k \times k$  matrix  $\mathbf{P}$ , where  $\mathbf{P}_{ij} = \langle w_i, w_j \rangle_{L^2(\nu)} = \int_0^\tau \omega_i(F(x))\omega_j(F(x))d\nu(x)$ , and then we set  $\tilde{\omega}_i(\cdot) = \sum_j \mathbf{P}_{ij}^{-1}w_j(\cdot)$ . Notice that the matrix  $\mathbf{P}$  may not have an inverse, meaning that the functions  $\omega_i$  are not linearly independent in  $L^2(\nu)$ . In such a case we compute its Moore-Penrose pseudo-inverse. Since, in practice, we do not have enough information to compute  $\mathbf{P}$ , we can estimate it (under the null hypothesis) by  $\hat{\mathbf{P}}_{ij} = n/(n_0n_1) \int_0^\tau \omega_i(\hat{F}(x-))\omega_j(\hat{F}(x-))L(x)d\hat{\Lambda}(x)$ . Finally, observe that this procedure generates a random kernel that converges to a deterministic one when  $n$  tends to infinity.

## 4 Asymptotic Properties

In this section we study some asymptotic properties of our test statistic  $Z_n$ . Before introducing our main results, we state some standard and very reasonable conditions that are needed in our proofs. We first state some smoothness, and moment conditions regarding the kernel  $K$ .

**Condition 3.** Let  $K : [0, 1]^2 \rightarrow \mathbb{R}$  be a continuous kernel, and let  $X, X', Y, Y'$  be independent random variables such that  $X, X' \sim F_0$  and  $Y, Y' \sim F_1$ . Then, we assume that

- i)  $\mathbb{E} (K(F(X), F(Y))^2(1 - G_0(X))(1 - G_1(Y))) < \infty$ ,
- ii)  $\mathbb{E} (K(F(X), F(X'))^2(1 - G_0(X))(1 - G_0(X'))) < \infty$ ,
- iii)  $\mathbb{E} (K(F(X), F(X))(1 - G_0(X))) < \infty$ ,
- iv)  $\mathbb{E} (K(F(Y), F(Y'))^2(1 - G_1(Y))(1 - G_1(Y'))) < \infty$ , and
- v)  $\mathbb{E} (K(F(Y), F(Y))(1 - G_1(Y))) < \infty$ .

In the previous conditions, recall that  $F = \eta F_0 + (1 - \eta)F_1$  is the pooled distribution, and  $F = F_0$  under the null hypothesis. In addition, we need a technical condition to deal with the randomness generated by  $\hat{F}(x-)$ , which is an analogue of the conditions needed in Theorem 4.2.1 of Gill [1980].

**Condition 4.** Assume that for every  $\varepsilon > 0$ , it holds that

- i)  $\lim_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_t^\tau K(\hat{F}(x-), \hat{F}(x-))(1 - G_0(x))dF_0(x) > \varepsilon \right) = 0$ , and
- ii)  $\lim_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_t^\tau K(\hat{F}(x-), \hat{F}(x-))(1 - G_1(x))dF_1(x) > \varepsilon \right) = 0$ .

Note that our conditions are not easy to verify in practice as they require knowledge of the distributions involved, however, the conditions are trivially satisfied for continuous and bounded kernels on  $[0, 1]^2$ , which is a property that most standard kernels, such as the Gaussian kernel, satisfy.

Finally, in order to ease notation, we present our results in terms of the functions  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$  and  $\psi^* : \mathbb{R}_+ \rightarrow \mathbb{R}$ , defined by

$$\psi(x) = \frac{(1 - G_0(x))(1 - G_1(x))}{\eta(1 - G_0(x)) + (1 - \eta)(1 - G_1(x))}, \quad (14)$$

and

$$\psi^*(x) = \frac{(1 - G_0(x))(1 - G_1(x))}{\eta(1 - H_0(x)) + (1 - \eta)(1 - H_1(x))}. \quad (15)$$

### 4.1 Limit distribution under the null hypothesis

Our first result establishes that, under the null hypothesis,  $(n_0n_1/n)Z_n$  converges in distribution to a limit random variable  $Z$ , when the number of data points tends to infinity.

**Theorem 5** (Limit distribution under the null). *Assume Conditions 3 and 4 hold. Then, under the null hypothesis, we have that*

$$\frac{n_0n_1}{n}Z_n \xrightarrow{\mathcal{D}} Z = \int_0^\tau K(F_0(x), F_0(x))\psi(x)dF_0(x) + \frac{1}{\eta(1 - \eta)}Y, \quad (16)$$

as  $n$  approaches infinity, where  $Y = \sum_{i=1}^\infty \lambda_i(\xi_i^2 - 1)$ ,  $\xi_1, \xi_2, \dots$  are a collection of (potentially infinity) i.i.d. standard normal random variables, and  $\lambda_1, \lambda_2, \dots$  are non-negative constants. Additionally, the mean and variance of  $Z$  are respectively given by

$$\mathbb{E}(Z) = \int_0^\tau K(F_0(x), F_0(x))\psi(x)dF_0(x) \quad \text{and} \quad \text{Var}(Z) = 2 \int_0^\tau \int_0^\tau K(F_0(x), F_0(y))^2\psi(x)\psi(y)dF_0(x)dF_0(y).$$

In the previous theorem, the constants  $\lambda_i$  are the eigenvalues of an integral operator in the  $L^2$  space associated with the measure induced by the triple  $(T_i, \Delta_i, c_i)$  under the random covariates model. More details about the limit distribution are shown in Section B along with the proof of Theorem 5.

## 4.2 Power under alternatives

We continue by studying the asymptotic behaviour of our test statistic under the alternative hypothesis, that is, under the assumption that  $H_1 : F_0 \neq F_1$  holds. To this end, we first prove that  $Z_n$  (without considering any scaling) converges to a deterministic value.

Define the measures  $\nu_0$  and  $\nu_1$  on  $\mathbb{R}_+$  by

$$\nu_0(A) = \int_A \psi^*(s) S_1(s) dF_0(s) \quad \text{and} \quad \nu_1(A) = \int_A \psi^*(s) S_0(s) dF_1(s), \quad (17)$$

where  $\psi^*$  is defined in Equation (15), and define their embeddings into  $\mathcal{H}$  by

$$\phi_0(\cdot) = \int_0^\tau K(F(y), \cdot) d\nu_0(y) \quad \text{and} \quad \phi_1(\cdot) = \int_0^\tau K(F(y), \cdot) d\nu_1(y), \quad (18)$$

respectively. The measures  $\nu_0$  and  $\nu_1$  can be understood as the population measures, respectively, of the empirical measures  $\nu_0^n$  and  $\nu_1^n$  defined in Equation (11). We will show that, under appropriate conditions, the embeddings  $\phi_0^n$  and  $\phi_1^n$ , defined in Equation (10), converge to  $\phi_0$  and  $\phi_1$ , respectively, where convergence is with respect to the norm of  $\mathcal{H}$ . This fact together with Theorem 1 yields the following result:

**Theorem 6.** *Assume Conditions 3 and 4 hold, then*

$$Z_n \xrightarrow{\mathbb{P}} \|\phi_0 - \phi_1\|_{\mathcal{H}}^2. \quad (19)$$

At this point, it should be clear that our test is consistent under the alternative hypothesis if  $\phi_0 \neq \phi_1$  whenever  $\nu_0 \neq \nu_1$ . We will show that if  $F_0 \neq F_1$  then  $\nu_0 \neq \nu_1$ , however, this result does not immediately extrapolate to  $\phi_0$  and  $\phi_1$ . A sufficient condition to ensure that  $\phi_0$  and  $\phi_1$  are different when  $F_0 \neq F_1$  is that the kernel  $K$  is  $c$ -universal (see definition in Section 3.1).

**Corollary 7** (Consistency). *Suppose that Conditions 3 and 4 hold, and that the kernel  $K$  is  $c$ -universal. Additionally, assume that  $1 - G_c(x) = 0$  implies  $S_c(x) = 0$ , for any  $c \in \{0, 1\}$ . Then, under the alternative hypothesis, we have that  $n_0 n_1 / n Z_n \rightarrow \infty$ , which implies that*

$$\Pr\left(\frac{n_0 n_1}{n} Z_n > Q_n(1 - \alpha)\right) \rightarrow 1,$$

where  $Q_n(1 - \alpha)$  denotes the  $(1 - \alpha)$ -quantile of the random variable  $(n_0 n_1 / n) Z_n$  under the null hypothesis.

We remark that the previous result uses the  $(1 - \alpha)$ -quantile of  $(n_0 n_1 / n) Z_n$  under the null hypothesis, which converges to a finite quantity due to Theorem 5.

Notice that Corollary 7 establishes consistency for all the alternatives under the assumption of a  $c$ -universal kernel  $K$ , nevertheless, even if the kernel  $K$  is not  $c$ -universal, we can still ensure consistency for particular alternatives, as we show in the following proposition.

**Proposition 8.** *Assume that Conditions 3 and 4 hold, and suppose that  $d\Lambda_1(x) = e^{\theta \bar{\omega}(F_0(x))} d\Lambda_0(x)$  for some  $\bar{\omega} \in \mathcal{H}$  and  $\theta \neq 0$ . Then  $(n_0 n_1 / n) Z_n \rightarrow \infty$ , and thus the test is consistent for such an alternative.*

Note that if  $d\Lambda_1(x) = e^{\theta \bar{\omega}(F_0(x))} d\Lambda_0(x)$  for some  $\bar{\omega} \in \mathcal{H}$  and  $\theta \neq 0$ , then  $\liminf_{n \rightarrow \infty} \text{LR}_n(\bar{\omega})^2 > 0$ . In this case,

$$\liminf_{n \rightarrow \infty} Z_n = \liminf_{n \rightarrow \infty} \sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}} \leq 1} \text{LR}_n(\omega)^2 \geq \liminf_{n \rightarrow \infty} \text{LR}_n(\bar{\omega})^2 > 0.$$

Therefore  $(n_0 n_1 / n) Z_n \rightarrow \infty$ , and thus the test is consistent for such a particular alternative. In general, any argument that ensures consistency of  $\text{LR}_n(\bar{\omega})^2$  for some  $\bar{\omega}$  in  $\mathcal{H}$  also applies to our test statistic  $Z_n$ , as long as the limit distribution of  $(n_0 n_1 / n) Z_n$  exists under the null hypothesis.

## 5 Wild Bootstrap Implementation

Given  $\alpha \in (0, 1)$ , we construct a statistical test of level  $\alpha$  by rejecting the null hypothesis if  $(n_0 n_1 / n) Z_n > Q_n(1 - \alpha)$ , where  $Q_n(1 - \alpha)$  is the  $(1 - \alpha)$ -quantile of the distribution of  $(n_0 n_1 / n) Z_n$  under the null hypothesis. As the distribution of  $(n_0 n_1 / n) Z_n$  is usually unknown, we can use its asymptotic distribution, given in Theorem 5, to approximate the rejection region. Unfortunately, excluding exceptional cases, the limiting distribution is rather complex, and thus computing the asymptotic  $(1 - \alpha)$ -quantile is very hard. In order to carry out our test, we introduce a simple Wild Bootstrap implementation of our testing procedure.

Recall, from Equation (8), that our test statistic  $Z_n$  can be written as follows:

$$Z_n = \left(\frac{n}{n_0 n_1}\right)^2 \sum_{i=1}^n \sum_{j=1}^n K(\hat{F}(X_i -), \hat{F}(X_j -)) L(X_i) L(X_j) (-1)^{c_i + c_j} \frac{\Delta_i}{Y_{c_i}(X_i)} \frac{\Delta_j}{Y_{c_j}(X_j)}. \quad (20)$$

Notice that the expression given for  $Z_n$  looks like a V-statistic. However, what prevents  $Z_n$  from being a V-statistic is that it depends on the functions  $\widehat{F}(x-)$ ,  $Y_0(x)/n_0$  and  $Y_1(x)/n_1$ , and  $n/(n_0n_1)L(x)$ , which are all random functions that depend on all data points. In Appendix B.1 and Appendix B.3, we prove that we can replace these functions by their respective limits plus some small additive error term that vanishes sufficiently fast as the number of observations grows. From this result, we deduce that  $Z_n$  can be approximated by a V-statistic, suggesting that the standard Wild Bootstrap sampling scheme of Dehling and Mikosch [1994] can be used to approximate the asymptotic distribution of  $n/(n_0n_1)Z_n$ .

Let  $\mathcal{W} = (W_i)_{i \in [n]}$  be a sequence of i.i.d. random variables satisfying that  $\mathbb{E}(W_i) = 0$  and  $\mathbb{E}(W_i^2) = 1$ , and such that they are independent of any other source of randomness, in particular, from the observations. Then, we define the Wild Bootstrap statistic associated with  $Z_n$  by

$$Z_n^{\mathcal{W}} = \left( \frac{n}{n_0n_1} \right)^2 \sum_{i=1}^n \sum_{j=1}^n W_i W_j K(\widehat{F}(X_i-), \widehat{F}(X_j-)) L(X_i) L(X_j) (-1)^{c_i+c_j} \frac{\Delta_i}{Y_{c_i}(X_i)} \frac{\Delta_j}{Y_{c_j}(X_j)}. \quad (21)$$

Similarly to the expression given in Equation (9), we can write  $Z_n^{\mathcal{W}}$  as

$$Z_n^{\mathcal{W}} = \left( \frac{n}{n_0n_1} \right)^2 (\mathbf{W}\mathbf{V})^\top \widehat{\mathbf{K}}(\mathbf{W}\mathbf{V}), \quad (22)$$

where  $\mathbf{W} \in \mathbb{R}^n \times \mathbb{R}^n$  is the diagonal matrix whose entries are given by  $\mathbf{W}_{jj} = W_j$ . Also, recall that  $\mathbf{V} \in \mathbb{R}^n$  is defined by  $\mathbf{V}_j = L(X_j)(-1)^{c_j} \Delta_j / Y_{c_j}(X_j)$  and  $\widehat{\mathbf{K}} \in \mathbb{R}^n \times \mathbb{R}^n$  is defined by  $\widehat{\mathbf{K}}_{ij} = K(\widehat{F}(X_i), \widehat{F}(X_j))$ . This matrix expression is very convenient for the computational implementation of our method.

The following theorem states the correctness of our Wild Bootstrap approach:

**Theorem 9.** *Assume that Conditions 3 and 4 are satisfied. Then, under the null hypothesis, it holds that for any  $x \in \mathbb{R}_+$ ,*

$$\mathbb{P} \left( \frac{n_0n_1}{n} Z_n^{\mathcal{W}} \geq x \mid (X_i, \Delta_i, c_i)_{i \in [n]} \right) \xrightarrow{\mathbb{P}} \mathbb{P}(Z \geq x),$$

where  $Z$  is the random variable defined in Theorem 5.

With the previous ingredients we are ready to describe our testing procedure, which relies on approximating the  $(1 - \alpha)$ -quantile of the distribution of  $Z_n$  by the quantiles of  $Z_n^{\mathcal{W}}$ . Since we can freely sample independent copies from  $Z_n^{\mathcal{W}}$  given the data points, we can estimate the quantile by Monte Carlo simulations. Our algorithm is as follows:

- i) Set level  $\alpha \in (0, 1)$  of the test and let  $N$  be a large integer,
- ii) Sample  $N$  independent copies of the Wild Bootstrap statistic using Equation (22),
- iii) Compute the  $(1 - \alpha)$ -quantile of the previous sample and call it  $Q_n^{\mathcal{W}}(1 - \alpha)$ ,
- iv) Compute the test statistic  $Z_n$  using Equation (9),
- v) Reject the null-hypothesis if  $Z_n > Q_n^{\mathcal{W}}(1 - \alpha)$ , otherwise do not reject it.

## 6 Simulations

We perform an empirical evaluation of our methods in which the ground truth is known. To this end, we consider two different settings: a *proportional hazard functions* setting (in which the classic log-rank test is probably the most powerful), and a *time-dependent hazard functions* setting, including Weibull and periodic hazard functions. All our experiments consider the same null cumulative hazard function  $\Lambda_0(t) = t$ , corresponding to the cumulative hazard function of an exponential random variable with mean 1. We choose  $\Lambda_1(t)$  belonging to one of the following parametric families:

- i) **Proportional hazards:** for this case we consider  $\Lambda_1(t)$  belonging to the parametric family given by  $\Lambda(t; \theta) = \theta t$  with  $\theta \in [0, 2]$ . Observe that  $\Lambda_0(t)$  is recovered when  $\theta = 1$ .
- ii) **Weibull (polynomial) hazards:** we consider  $\Lambda_1(t)$  belonging to a family of Weibull cumulative hazard functions  $\Lambda(t; \theta) = t^\theta$  with  $\theta \in [0, 2]$ . Notice that  $\theta = 1$  recovers the null.
- iii) **Periodic hazards:** we consider  $\Lambda_1(t)$  belonging to the family given by  $\Lambda(t; \theta) = t - \sin(\pi\theta)/(\pi\theta)$  with  $\theta \in (0, 15]$ . Notice that  $\lim_{\theta \rightarrow \infty} \Lambda(t; \theta) = \Lambda_0(t) = t$ , recovering the null hypothesis.

Figure 1 shows the behaviour of the cumulative hazard functions of the parametric families previously described, for different values of  $\theta$ .

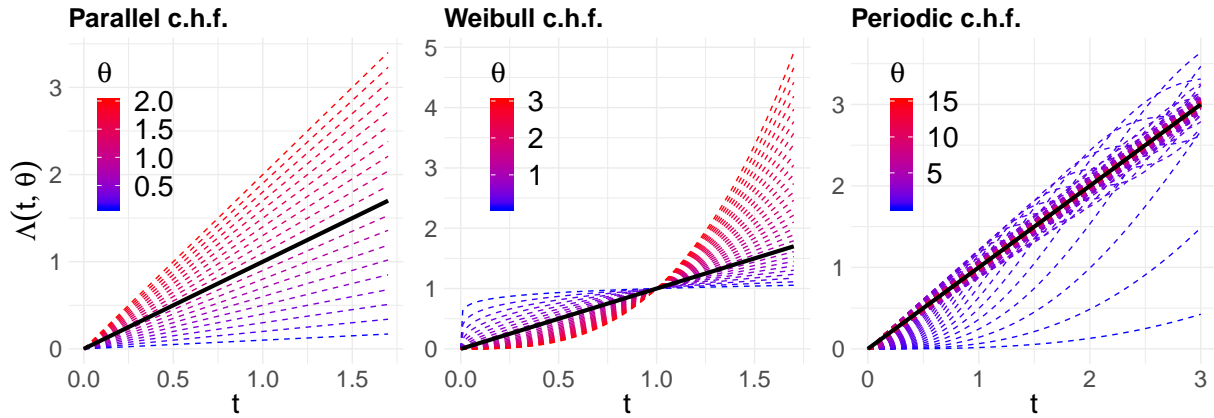


Figure 1: Set of cumulative hazard functions for the proportional, Weibull and periodic experiments. Solid black line denotes the null cumulative hazard  $\Lambda_0(t) = t$ . Observe that each family recovers the null cumulative hazard function  $\Lambda_0(t)$  for some specific value  $\theta_0$ .

## 6.1 Implementation

### 6.1.1 Kernels

The heart of our testing-approach is undoubtedly the kernel function. In Section 3.3 we showed that specific choices of the kernel function lead to some existing tests. In our experiments we use some of these choices as well as a kernel that is  $c$ -universal. We define kernels in the following categories:

1. **Log-rank kernels (LRP and LRC):** In Section 3.3.1, we considered kernels of the form  $K(x, y) = \omega(x)\omega(y)$ , recovering the well-known weighted log-rank tests. In our experiment we choose  $\omega$  to be equal to i)  $\omega_1(x) = 1$  and ii)  $\omega_2(x) = (x - 1/2)$ . For i) we recover the classical log-rank test (LRP), used to test proportional hazard functions, while for ii) we recover a weighted log-rank test (LRC), designed to discover if the hazard functions cross around the median of the distribution  $F$ .
2. **Projection kernels (P2W and P4W):** we follow the approach described in Section 3.3.3, which recovers the testing procedure of Brendel et al. [2014]. In particular, we choose kernel functions based on the subspace generated by the weight functions i)  $\{1, x\}$  and ii)  $\{1, x, x^2, x^3\}$ . We denote the kernels in i) and ii) by P2W and P4W, respectively, making a clear reference to the dimension of the subspace.
3. **Pearson-type kernels (Per4 and Per5):** we consider the Pearson-type kernel functions defined in Section 3.3.2, particularly in Equation (13). For our experiments, we partition the space into 4 and 5 disjoint regions, and denote the kernels by Per4 and Per5, respectively.
4. **Squared exponential kernel (SEK):** we consider the squared exponential kernel (SEK) defined by  $K(x, y) = \exp\{-(x - y)^2/\sigma^2\}$  with  $\sigma = 0.1$ . This kernel is  $c$ -universal, hence by Corollary 7, it leads to an omnibus test. Better results may be obtained by optimising the parameter  $\sigma$  for the problem at hand. A well-known heuristic is to choose  $\sigma$  as the median of the pairwise differences of the observations Scholkopf and Smola [2001].

### 6.1.2 Computer Implementation

Our experiments are implemented in R following the Wild Bootstrap approach described in Section 5, choosing Rademacher random variables for the Wild Bootstrap weights  $\mathcal{W}$ . We compute our test statistic  $Z_n$  and the Wild Bootstrap statistic  $Z_n^{\mathcal{W}}$  by using the quadratic form expressions given in Equations (9) and (22), respectively. By using these expressions, it is possible to give a simple and fast implementation of our testing procedure, indeed, 1000 repetitions of our testing procedure (using 1000 Wild Bootstrap samples) takes just a couple of minutes in a standard commercial laptop, for  $n_0 = n_1 = 100$ . For each experiment, we consider sample sizes of 30 and 100 observations per group, and we choose a censoring distribution generating 10% and 30% of censored observations.

## 6.2 Type-I error

In our first experiment, we verify that our tests achieve a correct Type-I error of  $\alpha = 0.05$  for each combination of sample size and censoring percentage, for the kernels described in Section 6.1.1. For each different combination of parameters, we run our test in 1000 simulated datasets. Table 1 shows the results. In general,

				Log-Rank		Projection		Pearson-type		SEK
$n_0$	$n_1$	cen <sub>0</sub>	cen <sub>1</sub>	LRP	LRC	P2W	P4W	Per4	Per5	SEK
30	30	10%	10%	4.5	5	5.1	4.7	<b>6.1</b>	<b>5.7</b>	5
		10%	30%	5.4	4.4	4.5	4.1	5	5.1	4.2
		30%	30%	4.3	4.7	4.2	4.4	4.7	4.4	4.2
30	100	10%	10%	3.9	<b>5.7</b>	5	5.1	5.2	4.9	5.2
		10%	30%	<b>6.7</b>	4.6	<b>7.1</b>	5.1	5.2	4.5	4.7
		30%	10%	5	<b>7.3</b>	<b>6.1</b>	4.9	<b>5.9</b>	<b>5.6</b>	5
		30%	30%	3.9	5.1	3.8	3.9	4.4	4.8	3.8
100	100	10%	10%	4.7	4.2	5.2	<b>5.9</b>	4.9	4.7	4.3
		10%	30%	5	<b>5.9</b>	5.4	<b>6.6</b>	<b>6.1</b>	<b>5.8</b>	<b>6.1</b>
		30%	30%	5.1	5.3	5.5	4.7	5.1	4.8	4.8

Table 1: Significance Level at  $\alpha = 5\%$ .  $n_c$  stands for the number of data points of group  $c$ , and cen $_c$  indicates the percentage of censored data.

the Wild Bootstrap approach has no problem reaching the correct level for even censoring percentages. Uneven censoring causes a few problems, so the user should be careful when applying this method in the latter setting. Arguably, the squared exponential kernel (SEK) is the most robust. It is worth recalling that projection tests use a random kernel (it depends on the data points) which might impact its performance. Additionally, while not reported here, we tested other significance levels, obtaining similar results.

### 6.3 Power Simulations

We provide an empirical evaluation of our testing procedure for each of the settings described in Section 6: proportional, Weibull and periodic hazard functions. As previously mentioned, the null cumulative hazard function is given by  $\Lambda_0(t) = t$ . The power is estimated by repeating our testing procedure over 1000 simulated datasets for each combination of sample size and censoring percentage. Results for the proportional, Weibull and periodic settings are shown in Figures 2, 3, and 4, respectively. We give a few comments and remarks about our experiments.

1. We report a very small fraction of our experiments as all the results are qualitatively the same as the ones shown in Figures 2, 3, and 4.
2. We only report results for the kernel Per5 since the kernel Per4 has an almost identical behaviour.
3. The LRP test is equivalent to the score test for the proportional hazard functions model; thus, it can be deduced that it is the most powerful test for local alternatives under this model. The previous statement is supported by Figure 2, where we observe an excellent performance of the LRP test in the setting of proportional hazard functions alternatives. Observe that the LRP test loses all of its power (nearly zero power) for the Weibull and periodic hazard alternatives as shown in Figures 3, and 4.
4. The LRC test is equivalent to the weighted log-rank test with weight  $w(x) = x - 1/2$ . It can be observed in Figure 3, that the LRC test has a very good performance in the setting of Weibull hazard alternatives, but its power is relatively low in the other two settings, as shown in Figures 2 and 4. An explanation for this behaviour follows from the fact that the LRC kernel is designed to be optimal at detecting a cross around the median of the pooled distribution  $F$ , and thus, it will not be a good fit for the proportional hazards experiment. Also, for more complex hazard functions, such as those described in the periodic setting, we can observe more than one cross occurring, which explains the poor performance of the LRC test.
5. We observe different behaviours for the projection kernels P2W and P4W. In the proportional hazard functions setting, the P2W test has the best performance after the LRP test, which is explained by the fact that this kernel is constructed considering the subspace generated by  $\{1, x\}$ , where the weight  $\omega(x) = 1$  is known to be optimal for proportional hazard alternatives. While P4W also includes this weight (it is generated by  $\{1, x, x^2, x^3\}$ ), the fact that it considers a larger space of possible alternatives makes the test more data-expensive resulting in a loss of power. For the Weibull hazard functions setting, we observe that both projection tests, P2W and P4W, have an overall good performance. This behaviour can be explained by the fact that both tests consider projections on polynomials, and the Weibull hazard functions are, indeed, polynomials. In the periodic case, both kernels have a substandard behaviour due to the fact that the hazard structure is rather different to a polynomial of finite degree. Note that with more data it seems that the tests do not improve when compared to the best kernel, in this case, the squared exponential kernel SEK.
6. The Pearson-type kernel Per5 has a consistent behaviour, being neither too good nor too bad. Disadvantages are that the user needs to specify beforehand a partition of the space.

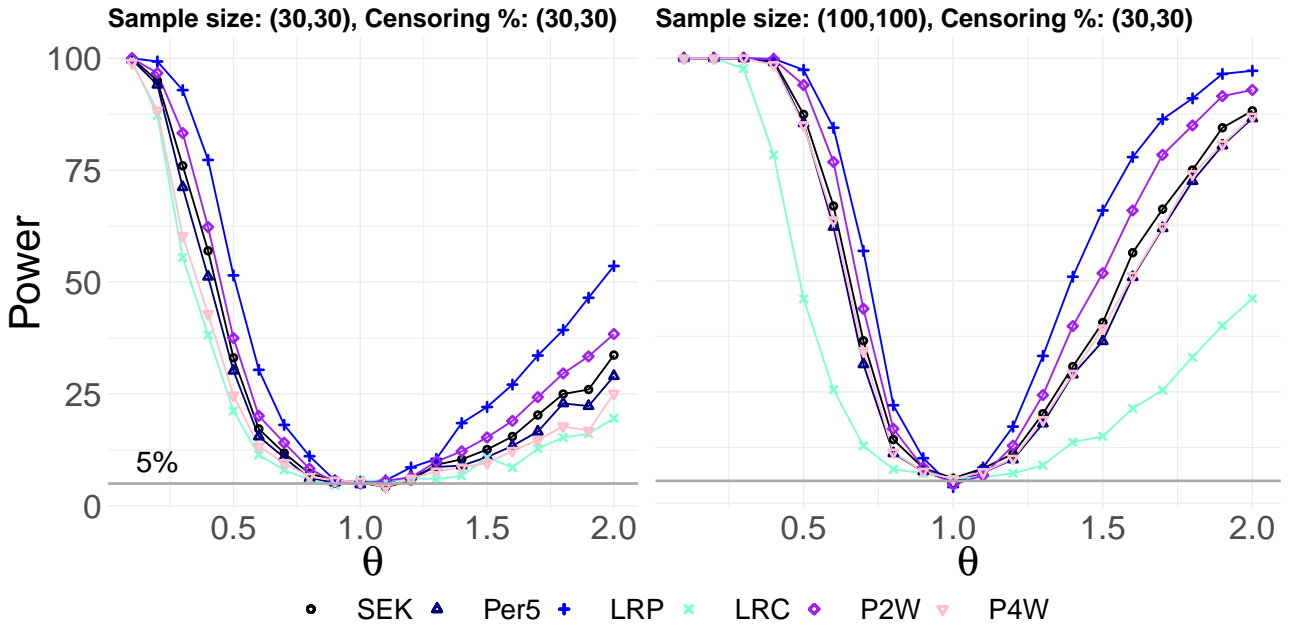


Figure 2: Proportional hazards experiment. Left: group sizes 30 and censoring percentage 30%. Right group sizes 100 and censoring percentage 30%.

	SEK	Per4	Per5	LRP	LRC	P2W	P4W
p-values	0.0053	0.0151	0.0222	0.2531	0.0011	0.0051	0.0228

Table 2:  $p$ -values obtained (using the Wild Bootstrap approach) for the comparison of the treatments ‘Chemotherapy’ and ‘Chemotherapy+Radiation’ for the GTSG data.

- The squared exponential kernel, SEK, gives overall good results. In one hand, while in the presence of more ‘structured’ data, i.e., proportional or Weibull (polynomial) hazard functions, simpler kernels give better results, the SEK still has a good performance. On the other hand, in presence of more ‘complex’ data, its performance is better than other kernels. In general, the SEK performs better than Pearson-type kernels, suggesting that the SEK is a better choice for an all-around kernel.
- In general, it seems that for nice structured data, simpler kernels (leading to simpler methods) have better results. On the other hand, for complex data, a more complex kernel seems to be a good option. Overall, we think that the SEK is the best option as it has a robust behaviour in simple data-scenarios, and it outperforms other tests in more complex scenarios. Also, as shown in Table 1, this kernel has close to no problems reaching the correct Type-I error.

## 7 Real data

We consider the Gastrointestinal Tumor Study Group data (GTSG), [Stablein et al. \[1981\]](#), available in the ‘`coin`’ R-package. The data considers a randomised clinical trial in the treatment of locally advanced, non-resectable gastric carcinoma. In this study, 42 patients were treated by using chemotherapy alone, while 45 patients were treated by using a combination of chemotherapy and radiation therapy. The aim of the study is to detect differences between these treatments. Kaplan-Meier curves for each group are shown in Figure 5. The null hypothesis is that there is no difference between the treatments. We apply our test considering the 7 different kernels described in Section 6.1.1. The corresponding  $p$ -values (approximated by using our Wild Bootstrap approach) are shown in Table 2.

All the tests, except for the classical log-rank test, reject the null hypothesis at 5% level of significance. This outcome is quite reasonable as the survival functions cross, as shown in Figure 5. Indeed, the smallest  $p$ -values are given by the SEK, LRC and the P2W tests which reject the null hypothesis at 1% level of significance. This is not a surprising behaviour of LRC and P2W, as these kernels are tailored to detect crossings. The SEK also performs very well which is quite satisfying as this kernel is not particularly designed for the setting of crossing hazards.

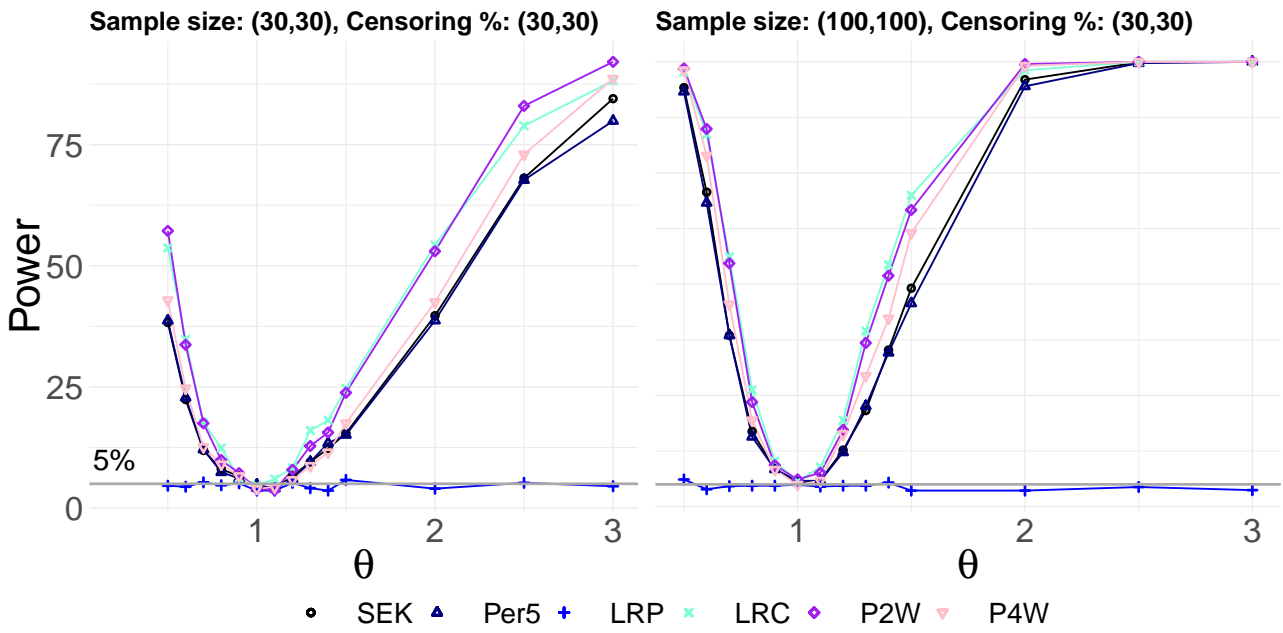


Figure 3: Polynomial (Weibull) hazards experiment. Left: group sizes 30 and censoring percentage 30%. Right group sizes 100 and censoring percentage 30%.

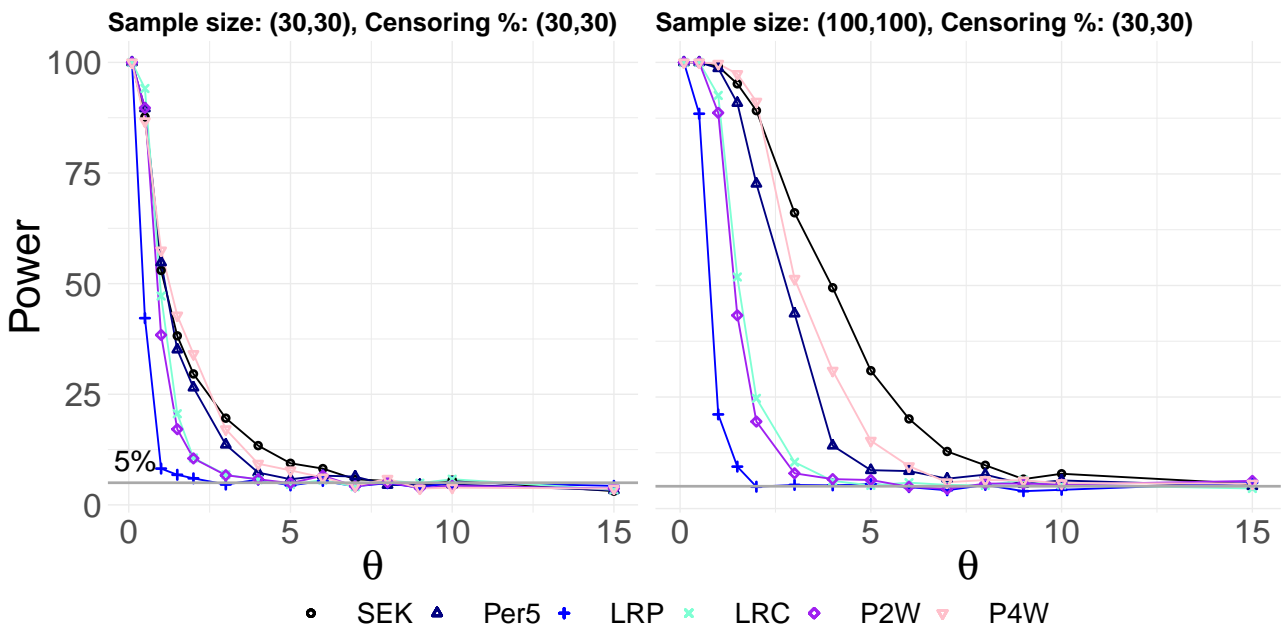


Figure 4: Periodic hazards experiment. Left: group sizes 30 and censoring percentage 30%. Right group sizes 100 and censoring percentage 30%.

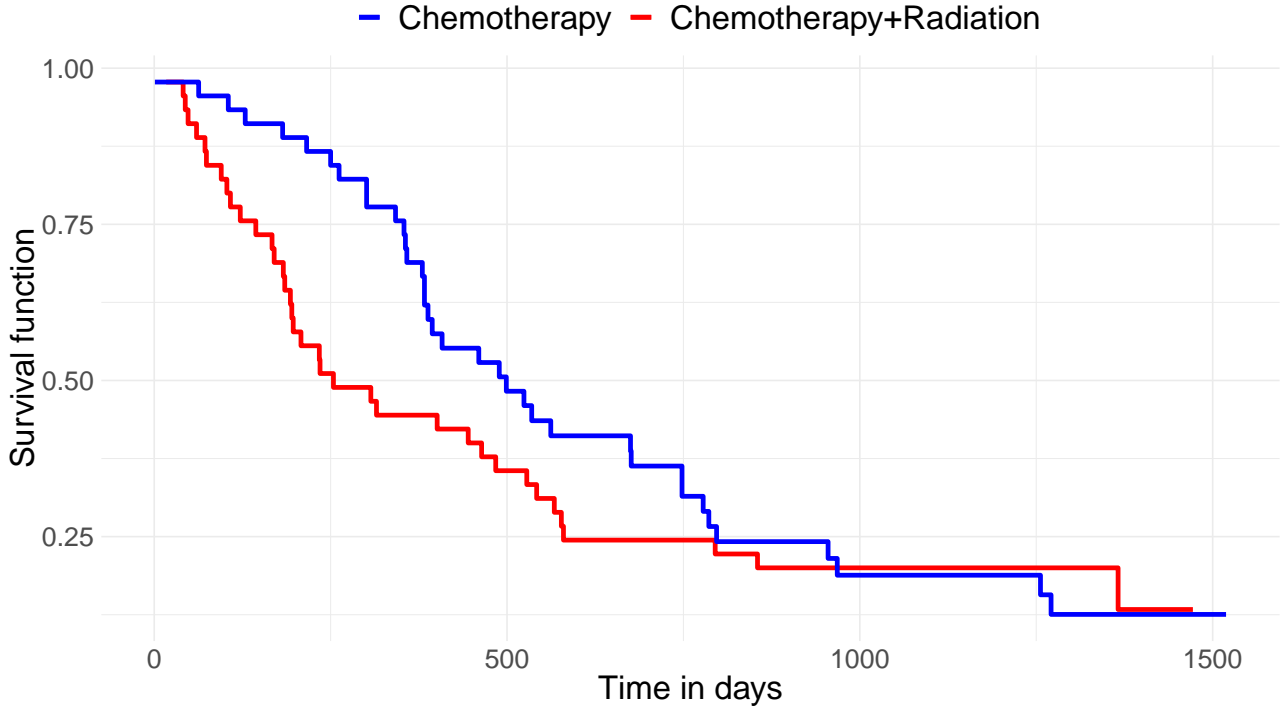


Figure 5: Kaplan-Meier curves for the patients treated by using chemotherapy alone versus a combination of chemotherapy and radiation therapy

## 8 Conclusion

We have introduced an RKHS-based testing procedure for the standard problem of two-sample hypothesis testing in the framework of right censored data. Our test statistic is the supremum of weighted log-rank statistics, where the weights belong to the unit ball of an RKHS. While our test statistic is apparently very complex, its evaluation becomes analytically tractable due to the reproducing property of reproducing kernel Hilbert spaces. Indeed, our test statistic can be written as a quadratic form as in Equation (9), and it can be fully characterised by a kernel function  $K$ . This simple structure allows us to derive asymptotic properties of our test, and it suggests that the standard Wild bootstrap approach can be used to approximate the rejection region. We also showed that our test can be seen as a natural infinitely-dimensional generalisation of other well-known test statistics based on log-rank statistics. Finally, we performed a simulation study which compares the results of the test statistic for different kernel functions.

To finish the paper we discuss some of our results and potential research ideas. First, as shown in Theorem (5), the asymptotic distribution of our test statistic under the null hypothesis is, in general, intractable, and thus we need to rely on bootstrapping techniques to approximate the rejection region. While, for us, the natural option is to consider a Wild Bootstrap approach, it is also possible to consider a permutation approach as the one used for weighted log-rank tests [Neuhauser, 1993]. In the case of equal censoring, the permutation approach has the advantage of being finitely exact, however, in the general case of censoring, the permutation approach is not directly applicable since the limit distribution of our test statistic depends on the censoring distribution of each group. For weighted log-rank tests, this problem is fixed by standardising the test statistic leading to an asymptotically distribution-free test. Unfortunately, it is not clear how to standardise (if possible) our test statistic, and we think this is a non-trivial task, especially for infinitely-dimensional kernels, hence, we leave this problem as future research.

A natural question to ask is: How do we choose a kernel function? Unfortunately, nobody can provide an answer for such a question yet and choosing kernel functions is an active field of research in Statistics and Machine Learning, as it is an issue that happens in several other contexts such as Gaussian Processes inference, support vector machine, kernel regression, etc. In practice, our simulation study suggests that simple (finite-dimensional) kernels, e.g., polynomials, are good enough if we know in advance that the hazard function has a relatively simple form, whereas complex kernels are better suited for complex hazard functions. In general, the squared exponential kernel is a very safe choice as it performs well in both settings. Nevertheless, we should not expect it to perform well in every setting, as Janssen [2000] proved that for a finite number of data points, any nonparametric test has preferences for a finite-dimensional subspace of



alternatives, and outside this subspace, the power of the test is almost flat in balls of alternatives. This means that, for a fixed number of data points, it is possible to find alternatives for which the squared exponential kernel fails, however, due to the complex form of the kernel, it may be difficult to construct such an alternative, and we do not expect that to happen in practical settings. Note that, by choosing a reproducing kernel, Theorem 5 shows that our test fixes most of its power on a finite number of directions, which are given by the eigenfunctions with larger eigenvalues in the spectral decomposition of the kernel, and thus, with a finite number of data points we expect the test to concentrate its power around the first  $m$  eigenfunctions with larger eigenvalues, where  $m$  is some constant depending on  $n$ , and as long as  $n$  grows,  $m$  should be growing as well.

Finally, we give a few comments about our technical results. Our asymptotic analysis uses the fact that our test statistic is a double stochastic integral (Equation (7)), and particularly, under the null hypothesis, those stochastic integrals are with respect to martingales. Our main tool to study these objects is Theorem 17, which allows us to control double stochastic integrals with respect to a special class of random integrands  $h$ , which includes our integrand  $K(\widehat{F}(x-), \widehat{F}(y-))L(x)L(y)$ , as well as several others. Our theorem is powerful enough to study other type of collection of log-rank statistics, for example, we can consider the test statistic

$$\sup_{\omega \in \mathcal{H}, \|\omega\|_{\mathcal{H}}^2 \leq 1} \int_0^{\tau_n} \omega(Y(x)/n)L(x)(d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x)),$$

which is similar to our test statistic (c.f., Equations (4) and (1)), but replacing  $\widehat{S}(x-)$  by  $Y(x)/n$  (See [Fleming and Harrington, 1991, Chapter 7]). In general, replacing  $\widehat{S}(x-)$  by any reasonable predictable process gives a test statistic that can be analysed by our methods, and furthermore, our tool can be applicable to an even larger class of processes. The analysis of other type of test statistics is subject of future research.

## Correspondence

Nicolas Rivera.

Address: Computer Laboratory, University of Cambridge. William Gates Building, 15 JJ Thomson Ave, Cambridge CB3 0FD, UK.

e-mail: nr454 at cam dot ac dot uk

## References

- Michael G. Akritas. Pearson-type goodness-of-fit tests: the univariate case. *J. Amer. Statist. Assoc.*, 83(401):222–230, 1988. ISSN 0162-1459. URL [http://links.jstor.org/sici?sici=0162-1459\(198803\)83:401<222:PGTUC>2.0.CO;2-0&origin=MSN](http://links.jstor.org/sici?sici=0162-1459(198803)83:401<222:PGTUC>2.0.CO;2-0&origin=MSN).
- Michael G. Akritas. The central limit theorem under censoring. *Bernoulli*, 6(6):1109–1120, 2000. ISSN 1350-7265. doi: 10.2307/3318473. URL <https://doi.org/10.2307/3318473>.
- Per K Andersen, Ornulf Borgan, Richard D Gill, and Niels Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.
- N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950. ISSN 0002-9947. doi: 10.2307/1990404. URL <https://doi.org/10.2307/1990404>.
- V Bagdonavicius, J Kruopis, and M S Nikulin. *Non-parametric tests for censored data*. John Wiley & Sons, 2010.
- Arne Bathke, Mi-Ok Kim, and Mai Zhou. Combined multiple testing by censored empirical likelihood. *J. Statist. Plann. Inference*, 139(3):814–827, 2009. ISSN 0378-3758. doi: 10.1016/j.jspi.2008.05.041. URL <https://doi.org/10.1016/j.jspi.2008.05.041>.
- Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers, Boston, MA, 2004. ISBN 1-4020-7679-7. doi: 10.1007/978-1-4419-9096-9. URL <https://doi.org/10.1007/978-1-4419-9096-9>. With a preface by Persi Diaconis.
- Arup Bose and Arusharka Sen. Asymptotic distribution of the Kaplan-Meier  $U$ -statistics. *J. Multivariate Anal.*, 83(1):84–123, 2002. ISSN 0047-259X. doi: 10.1006/jmva.2001.2039. URL <https://doi.org/10.1006/jmva.2001.2039>.
- Michael Brendel, Arnold Janssen, Claus-Dieter Mayer, and Markus Pauly. Weighted logrank permutation tests for randomly right censored life science data. *Scand. J. Stat.*, 41(3):742–761, 2014. ISSN 0303-6898. doi: 10.1111/sjos.12059. URL <https://doi.org/10.1111/sjos.12059>.

- Herold Dehling and Thomas Mikosch. Random quadratic forms and the bootstrap for  $U$ -statistics. *J. Multivariate Anal.*, 51(2):392–413, 1994. ISSN 0047-259X. doi: 10.1006/jmva.1994.1069. URL <https://doi.org/10.1006/jmva.1994.1069>.
- Marc Ditzhaus and Sarah Friedrich. More powerful logrank permutation tests for two-sample survival data, 2018.
- Bradley Efron and Iain M. Johnstone. Fisher’s information in terms of the hazard rate. *Ann. Statist.*, 18(1):38–62, 1990. ISSN 0090-5364. doi: 10.1214/aos/1176347492. URL <https://doi.org/10.1214/aos/1176347492>.
- Tamara Fernandez and Arthur Gretton. A maximum-mean-discrepancy goodness-of-fit test for censored data. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 2966–2975. PMLR, 2019. URL <http://proceedings.mlr.press/v89/fernandez19a.html>.
- Tamara Fernández and Nicolás Rivera. Kaplan-meier  $v$ - and  $u$ -statistics. *Electron. J. Statist.*, 14(1):1872–1916, 2020. ISSN 1935-7524. doi: 10.1214/20-EJS1704.
- Thomas R. Fleming and David P. Harrington. *Counting processes and survival analysis*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1991. ISBN 0-471-52218-X.
- Kenji Fukumizu, Francis R. Bach, and Michael I. Jordan. Kernel dimension reduction in regression. *Ann. Statist.*, 37(4):1871–1905, 2009. ISSN 0090-5364. doi: 10.1214/08-AOS637. URL <https://doi.org/10.1214/08-AOS637>.
- Valérie Garès, Sandrine Andrieu, Jean-François Dupuy, and Nicolas Savy. An omnibus test for several hazard alternatives in prevention randomized controlled clinical trials. *Statistics in medicine*, 34(4):541–557, 2015.
- Richard Gill et al. Large sample behaviour of the product-limit estimator on the whole line. *The Annals of Statistics*, 11(1):49–58, 1983.
- Richard D Gill. Censoring and stochastic integrals. *Statistica Neerlandica*, 34(2):124–124, 1980.
- Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.
- David P Harrington and Thomas R Fleming. A class of rank test procedures for censored survival data. *Biometrika*, 69(3):553–566, 1982.
- Svante Janson. Probability asymptotics: notes on notation, 2011.
- Arnold Janssen. Global power functions of goodness of fit tests. *Annals of Statistics*, pages 239–253, 2000.
- John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2006.
- V. S. Koroljuk and Yu. V. Borovskich. *Theory of  $U$ -statistics*, volume 273 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994. ISBN 0-7923-2608-3. doi: 10.1007/978-94-017-3515-5. URL <https://doi.org/10.1007/978-94-017-3515-5>.
- Michael R Kosorok and Chin-Yu Lin. The versatility of function-indexed weighted log-rank statistics. *Journal of the American Statistical Association*, 94(445):320–332, 1999.
- Wolfgang Kössler. Max-type rank tests,  $u$ -tests, and adaptive tests for the two-sample location problem an asymptotic power study. *Computational Statistics & Data Analysis*, 54(9):2053–2065, 2010.
- J. A. Koziol and Y. S. Yuh. Omnibus two-sample test procedures with randomly censored data. *Biometrical J.*, 24(8):743–750, 1982. ISSN 0323-3847. doi: 10.1002/bimj.4710240804. URL <https://doi.org/10.1002/bimj.4710240804>.
- James A. Koziol and Sylvan B. Green. A Cramér-von Mises statistic for randomly censored data. *Biometrika*, 63(3):465–474, 1976. ISSN 0006-3444. doi: 10.1093/biomet/63.3.465. URL <https://doi.org/10.1093/biomet/63.3.465>.
- Tze Leung Lai, Zhiliang Ying, et al. Rank regression methods for left-truncated and right-censored data. *The Annals of Statistics*, 19(2):531–556, 1991.

- N Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, 50(3):163, 1966.
- Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1-2): 1–141, 2017. ISSN 1935-8237. doi: 10.1561/22000000060. URL <http://dx.doi.org/10.1561/22000000060>.
- Georg Neuhaus. Local asymptotics for linear rank statistics with estimated score functions. *Ann. Statist.*, 15(2):491–512, 1987. ISSN 0090-5364. doi: 10.1214/aos/1176350357. URL <https://doi.org/10.1214/aos/1176350357>.
- Georg Neuhaus. Conditional rank tests for the two-sample problem under random censorship. *Ann. Statist.*, 21(4):1760–1779, 1993. ISSN 0090-5364. doi: 10.1214/aos/1176349396. URL <https://doi.org/10.1214/aos/1176349396>.
- Margaret Sullivan Pepe and Thomas R. Fleming. Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data. *Biometrics*, 45(2):497–507, 1989. ISSN 0006-341X. doi: 10.2307/2531492. URL <https://doi.org/10.2307/2531492>.
- Richard Peto and Julian Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2):185–207, 1972. ISSN 00359238. URL <http://www.jstor.org/stable/2344317>.
- Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001. ISBN 0262194759.
- Yu Shen and Jianwen Cai. Maximum of the weighted Kaplan-Meier tests with application to cancer prevention and screening trials. *Biometrics*, 57(3):837–843, 2001. ISSN 0006-341X. doi: 10.1111/j.0006-341X.2001.00837.x. URL <https://doi.org/10.1111/j.0006-341X.2001.00837.x>.
- Carl-Johann Simon-Gabriel and Bernhard Schölkopf. Kernel distribution embeddings: universal kernels, characteristic kernels and kernel metrics on distributions. *J. Mach. Learn. Res.*, 19:Paper No. 44, 29, 2018. ISSN 1532-4435.
- Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In *International Conference on Algorithmic Learning Theory*, pages 13–31. Springer, 2007.
- Cesar Souza. Kernel functions for machine learning applications, 2010. URL <http://crsouza.com/2010/03/17/kernel-functions-for-machine-learning-applications>.
- Donald M Stablein, Walter H Carter Jr, and Joel W Novak. Analysis of survival data with nonproportional hazard functions. *Controlled clinical trials*, 2(2):149–159, 1981.
- W. Stute and J.-L. Wang. The strong law under random censorship. *Ann. Statist.*, 21(3):1591–1607, 1993. ISSN 0090-5364. doi: 10.1214/aos/1176349273. URL <https://doi.org/10.1214/aos/1176349273>.
- Robert E. Tarone. On the distribution of the maximum of the logrank statistic and the modified wilcoxon statistic. *Biometrics*, 37(1):79–85, 1981. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/2530524>.
- Robert E Tarone and James Ware. On distribution-free tests for equality of survival distributions. *Biometrika*, 64(1):156–160, 1977.
- Song Yang and Ross Prentice. Improved logrank-type tests for survival data using adaptive weights. *Biometrics*, 66(1):30–38, 2010.
- Song Yang, Li Hsu, and Lueping Zhao. Combining asymptotically normal tests: case studies in comparison of two groups. *Journal of statistical planning and inference*, 133(1):139–158, 2005.
- Mai Zhou. *Empirical likelihood method in survival analysis*. Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL, 2016. ISBN 978-1-4665-5492-4.

# Appendix

## A Preliminary Results

### A.1 Some Results for Counting Processes

**Proposition 10.** *The following results hold:*

- i)  $\lim_{n \rightarrow \infty} \sup_{t \leq \tau} |\widehat{S}(t) - S(t)| = 0$  a.s.,
- ii)  $\lim_{n \rightarrow \infty} \sup_{t \leq \tau} |Y(t)/n - H(t-)| = 0$  a.s.,

Item **i** is due to [Stute and Wang \[1993\]](#), and item **ii** is the Glivenko-Cantelli theorem.

**Proposition 11.** *Let  $\beta \in (0, 1)$ , then*

- i)  $\mathbb{P}(\widehat{S}(t) \leq \beta^{-1}S(t), \quad \forall t \leq \tau_n) \geq 1 - \beta$ ,
  - ii)  $\mathbb{P}(Y(t)/n \leq \beta^{-1}(1 - H(t-)), \quad \forall t \leq \tau_n) \geq 1 - \beta$ , and
  - iii)  $\mathbb{P}(Y(t)/n \geq \beta(1 - H(t-)), \quad \forall t \leq \tau_n) \geq 1 - e(1/\beta)e^{-1/\beta}$ .
- i.e., i)  $\sup_{t \leq \tau_n} \widehat{S}(t)/S(t) = O_p(1)$ , ii) and iii)  $\sup_{t \leq \tau_n} Y(t)/(n(1 - H(t-))) = \Theta_p(1)$ .*

The proofs of items **i** and **iii** are due [Gill et al. \[1983\]](#), and item **ii** follows from [Gill \[1980, Theorem 3.2.1\]](#).

*Remark 12.* Notice that the results of the previous propositions still hold if we replace  $Y(t)$ ,  $H(t)$  and  $n$  by  $Y_c(t)$ ,  $H_c(t)$  and  $n_c$ , respectively.

### A.2 Double Martingales

In our proofs, we will frequently encounter double martingale integral processes of the form:

$$\int_0^t \int_0^t h_n(x, y) dW_n(x) dW_n(y), \quad t \geq 0, \quad (23)$$

where  $h_n : \mathbb{R}^2 \rightarrow \mathbb{R}$  is a sequence of symmetric positive-definite random functions, and  $(W_n(t))_{t \geq 0}$  is a sequence of  $(\mathcal{F}_t)$ -martingales. The aim of this section is to establish conditions under which such processes converge to zero in probability when evaluated at  $t = \tau_n$ . This result is formally stated in [Theorem 17](#), and to prove it, we use some results introduced by [Fernández and Rivera \[2020\]](#) for double integrals with respect to martingales.

**Definition 13.** Define the  $\sigma$ -algebra  $\mathcal{P}$  on  $C = \{(x, y) : 0 < x < y < \infty\}$  as the  $\sigma$ -algebra generated by sets of the form

$$(a_1, b_1] \times (a_2, b_2] \times X, \text{ where } 0 \leq a_1 \leq b_1 < a_2 \leq b_2, \text{ and } X \in \mathcal{F}_{a_2},$$

and  $\{0\} \times \{0\} \times X$ , where  $X \in \mathcal{F}_0$ .

A stochastic process  $(h(x, y))_{(x, y) \in C}$  is said to be  $\mathcal{P}$ -measurable if it is measurable with respect to  $\mathcal{P}$ .

The following proposition is a simple consequence of the definition of  $\mathcal{P}$ .

**Proposition 14.** *Let  $k : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a measurable function, and let  $(h_1(t))_{t \geq 0}$  and  $(h_2(t))_{t \geq 0}$  be  $(\mathcal{F}_t)$ -predictable stochastic processes. Then the process  $(h(x, y))_{(x, y) \in C}$  given by  $h(x, y) = k(h_1(x), h_2(y))$  is  $\mathcal{P}$ -measurable.*

**Theorem 15.** *Let  $h$  be a  $\mathcal{P}$ -measurable process, and let  $W$  be a right-continuous  $(\mathcal{F}_t)$ -martingale. Assume that for all  $t \geq 0$ , it holds that*

$$\mathbb{E} \left( \int_{(0, t]} \int_{(0, y)} |h(x, y)| |dW(x)| |dW(y)| \right) < \infty. \quad (24)$$

*Then,  $(Z(t))_{t \geq 0}$ , defined by  $Z(t) = \int_0^t \int_{(0, y)} h(x, y) dW(x) dW(y)$ , is an  $(\mathcal{F}_t)$ -martingale.*

In our proofs we are particularly interested in predictable positive-definite processes which are defined as following:

**Definition 16.** We say a process  $(h(x, y))_{x, y \geq 0}$  is a *predictable positive-definite* process if it satisfies the following properties: i)  $h(x, y) = h(y, x)$ , ii)  $h$  is positive definite, i.e., each realisation of the stochastic process is a positive definite function, iii)  $(h(x, y))_{(x, y) \in C}$  is  $\mathcal{P}$ -measurable, and iv)  $(h(x, x))_{x \geq 0}$  is predictable with respect to  $(\mathcal{F}_x)_{x \geq 0}$ .

The next theorem, whose proof is deferred to [Section E](#), gives sufficient conditions under which the process of [Equation \(23\)](#) converges to zero in probability.

**Theorem 17.** Let  $(h_n(x, y))_{x, y \geq 0}$  be a sequence of predictable positive-definite processes, let  $W_n$  be a sequence of right-continuous  $(\mathcal{F}_t)$ -martingales with predictable and quadratic variation processes denoted by  $\langle W_n \rangle$  and  $[W_n]$ , respectively, and suppose that

$$\mathbb{E} \left( \int_0^t \int_0^t |h_n(x, y)| |dW_n(x)| |dW_n(y)| \right) < \infty \quad (25)$$

holds for all  $n$  large enough, and for all  $t \in (0, \tau)$ .

Then, if

$$\int_0^{\tau_n} h_n(x, x) d\langle W_n \rangle(x) = o_p(1),$$

we have that

$$\int_0^{\tau_n} \int_0^{\tau_n} h_n(x, y) dW_n(x) dW_n(y) = o_p(1),$$

and the same holds if we replace  $o_p(1)$  by  $O_p(1)$ .

Note that Equation (25) holds trivially due to the simple nature of our martingales, hence, we will not verify this conditions in our applications of the theorem.

## B Analysis under the Null Hypothesis: Proof of Theorem 5

The proof of Theorem 5 is split into three mains steps:

- i) We find a cleaner asymptotic expression for our test statistic under the null hypothesis. In particular, we show that  $Z_n$  can be rewritten as

$$Z_n = \left( \frac{n}{n_0 n_1} \right)^2 \sum_{i=1}^n \sum_{j=1}^n \int_0^{\tau_n} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(y-)) \frac{(-1)^{c_i + c_j} L(x) L(y)}{Y_{c_i}(x) Y_{c_j}(y)} dM^i(x) dM^j(y).$$

Then, by using Theorem 17, we prove that  $\widehat{F}(x-)$ ,  $n/(n_0 n_1) L(x)$ , and  $Y_c(x)/n$  can be replaced by their respective limits,  $F_0(x)$ ,  $\psi(x) S_0(x)$ , and  $(1 - H_c)(x) \eta^{c-1} (1 - \eta)^{-c}$ , up to a small additive term that decreases to zero in probability, obtaining that

$$Z_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_0^{\tau_n} \int_0^{\tau_n} \frac{K(F_0(x), F_0(y)) \psi(x) \psi(y)}{(1 - G_{c_i}(x))(1 - G_{c_j}(y))} \frac{1}{\eta^2} \left( \frac{-\eta}{(1 - \eta)} \right)^{c_i + c_j} dM^i(x) dM^j(y) + o_p(n^{-1}). \quad (26)$$

- ii) We prove that the deterministic and random covariates models are asymptotically equivalent in the sense that our test statistic  $nZ_n$  has the same asymptotic distribution (when it exists) under both models.
- iii) We obtain the limit distribution of  $nZ_n$  under the random covariates model. Our results translate to the deterministic covariates model by using the result of the previous item.

### B.1 Step i: Finding a simpler asymptotic representation

For this step, we work under the deterministic covariates model, but notice that our analysis can be extended to the random covariates model by conditioning on the number of random covariates with value 0, say  $N_0$ , and by noticing that  $N_0/n \rightarrow \eta$  almost surely.

From Theorem 1 and Lemma 2, we have that

$$Z_n = \left\| \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 \quad (27)$$

$$= \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(y-)) L(x) L(y) d\Lambda^*(x) d\Lambda^*(y), \quad (28)$$

where  $d\Lambda^*(x) = d\widehat{\Lambda}_0(x) - d\widehat{\Lambda}_1(x)$ , and

$$d\Lambda^*(x) = \frac{dM_0(x)}{Y_0(x)} - \frac{dM_1(x)}{Y_1(x)} = \sum_{i=1}^n (-1)^{c_i} \frac{dM^i(x)}{Y_{c_i}(x)} \quad (29)$$

holds under the null hypothesis. In particular, notice that, under the null hypothesis,  $(\Lambda^*(x))_{x \geq 0}$  is an  $(\mathcal{F}_x)$ -martingale with predictable and quadratic variation processes given by

$$d\langle \Lambda^* \rangle(x) = \frac{d\Lambda_0(x)}{L(x)} \quad \text{and} \quad d[\Lambda^*](x) = \frac{dN_0(x)}{Y_0(x)^2} + \frac{dN_1(x)}{Y_1(x)^2}, \quad (30)$$

respectively. Then, by substituting Equation (29) in Equation (28), we can rewrite our test statistic as

$$Z_n = \left( \frac{n}{n_0 n_1} \right)^2 \sum_{i=1}^n \sum_{j=1}^n \int_0^{\tau_n} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(y-)) L(x) L(y) (-1)^{c_i + c_j} \frac{dM^i(x) dM^j(y)}{Y_{c_i}(x) Y_{c_j}(y)}.$$

The main result of this section is the following:

**Theorem 18.** *Assume Conditions 3 and 4 hold. Then, under the null hypothesis, it holds that*

$$nZ_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \int_0^{\tau_n} \int_0^{\tau_n} \frac{K(F_0(x), F_0(y)) \psi(x) \psi(y)}{(1 - G_{c_i}(x))(1 - G_{c_j}(y))} \eta^{-2} \left( \frac{-\eta}{(1 - \eta)} \right)^{c_i + c_j} dM^i(x) dM^j(y) + o_p(1).$$

We split the proof of Theorem 18 into three parts:

- 1) We prove that the pooled Kaplan-Meier estimator,  $\widehat{F}(x-)$  can be replaced by its limit  $F_0(x)$  (recall  $F_0$  is continuous, and thus  $F_0(x) = F_0(x-)$ ),
- 2) we prove that  $nL(x)/(n_0 n_1)$  can be replaced by its limit  $\psi(x)S_0(x)$ , and
- 3) we prove that  $Y_0(x)/n$  and  $Y_1(x)/n$  can be replaced by their limits,  $\eta(1 - H_0(x))$  and  $(1 - \eta)(1 - H_1(x))$ , respectively.

For the proof of Theorem 18 we will recurrently use the following fact: let  $q: \mathbb{R} \rightarrow [0, 1)$  and let  $\mu$  be a measure on  $[0, \infty)$ , then

$$\left\| \int_0^\infty K(q(x), \cdot) \mu(dx) \right\|_{\mathcal{H}}^2 = \int_0^\infty \int_0^\infty K(q(x), q(y)) \mu(dx) \mu(dy),$$

which is a straightforward consequence of the linearity of the inner product and integration, and the reproducing property (this argument was used in the proof of Theorem 1).

## B.2 Proof of Theorem 18

**Part 1): Replacement of  $\widehat{F}(x-)$  by  $F_0(x)$**

The following Lemma proves that  $K(\widehat{F}(x-), \widehat{F}(y-))$  can be replaced by  $K(F_0(x), F_0(y))$ , up to a small error.

**Lemma 19.** *Assume Conditions 3 and 4. Then, under the null hypothesis, it holds that*

$$nZ_n = n \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} K(F_0(x), F_0(y)) L(x) L(y) d\Lambda^*(x) d\Lambda^*(y) + o_p(1).$$

*Proof.* Using norm notation, the desired result is equivalent to show that

$$n \left\| \left( \frac{n}{n_0 n_1} \right) \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = n \left\| \left( \frac{n}{n_0 n_1} \right) \int_0^{\tau_n} K(F(x), \cdot) L(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 + o_p(1). \quad (31)$$

By triangular inequality  $\|b\|_{\mathcal{H}} - \|a - b\|_{\mathcal{H}} \leq \|a\|_{\mathcal{H}} \leq \|b\|_{\mathcal{H}} + \|a - b\|_{\mathcal{H}}$ , then, by taking square, we deduce that we just need to prove that

$$n \left\| \frac{n}{n_0 n_1} \int_0^{\tau_n} (K(F(x), \cdot) - K(\widehat{F}(x-), \cdot)) L(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = o_p(1), \quad (32)$$

and that

$$n \left\| \frac{n}{n_0 n_1} \int_0^{\tau_n} K(F(x), \cdot) L(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = O_p(1). \quad (33)$$

We begin by verifying Equation (33). Expanding the inner product expression, the left-hand side equals

$$O(1) \frac{1}{n} \int_0^{\tau_n} \int_0^{\tau_n} K(F(x), F(y)) L(x) L(y) d\Lambda^*(x) d\Lambda^*(y).$$

Also, notice that  $(K(F(x), F(y))L(x)L(y))_{x, y \geq 0}$  is a predictable positive-definite process (recall Definition 16), and since  $(\Lambda^*(x))_{x \geq 0}$  is an  $(\mathcal{F}_x)$ -martingale with predictable and quadratic variation processes given in Equation (30), a straightforward application of Theorem 17 tell us that we just need to verify that

$$\frac{1}{n} \int_0^{\tau_n} K(F(x), F(x)) L(x) d\Lambda_0(x) = O_p(1).$$

The previous equation holds true by Condition 3, and by using that  $L(x)/n = O_p(1)S_0(x)\psi(x)$  uniformly for all  $x \leq \tau_n$  due to Propositions 11.ii and 11.iii, and since  $\psi(x) \leq \eta^{-1}(1 - G_0(x))$ .

We continue verifying Equation (32). Notice that

$$n \left\| \frac{n}{n_0 n_1} \int_0^{\tau_n} (K(F(x), \cdot) - K(\widehat{F}(x-), \cdot))L(x)d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = n \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} h(x, y)d\Lambda^*(x)d\Lambda^*(y), \quad (34)$$

where

$$h(x, y) = \left( K(F_0(x), F_0(y)) - K(F_0(x), \widehat{F}(y-)) - K(\widehat{F}(x-), F_0(y)) + K(\widehat{F}(x-), \widehat{F}(y-)) \right) L(x)L(y).$$

It is easy to verify that  $(h(x, y))_{x, y \geq 0}$  is a predictable positive-definite process, then, by using Theorem 17, the desired result follows from proving  $n \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} h(x, x)d\Lambda^*(x) = o_p(1)$ .

Using that  $L(x)/n = O_p(1)S_0(x)\psi(x)$  uniformly for all  $x \leq \tau_n$  and  $\psi(x) \leq \eta^{-1}(1 - G_0(x))$ , we get

$$n \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} h(x, x)d\Lambda^*(x) = O_p(1) \int_0^{\tau} \omega(x)(1 - G_0(x))dF_0(x),$$

where  $\omega(x) = K(F_0(x), F_0(x)) - 2K(F_0(x), \widehat{F}(x-)) + K(\widehat{F}(x-), \widehat{F}(x-))$ . Let  $\varepsilon > 0$ , then for any  $t \in (0, \tau)$ ,

$$\begin{aligned} & \mathbb{P} \left( \int_0^{\tau} \omega(x)(1 - G_0(x))dF_0(x) \geq \varepsilon \right) \\ & \leq \mathbb{P} \left( \int_0^t \omega(x)(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{2} \right) + \mathbb{P} \left( \int_t^{\tau} \omega(x)(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{2} \right) \end{aligned} \quad (35)$$

We will prove that both terms on the right-hand side of the previous equation tends to 0 as  $n$  approaches infinity. For the first term, notice that, since  $\sup_{x \leq t} |\widehat{F}(x-) - F(x)| = 0$  a.s. by Proposition 10.i, and since  $K$  is continuous in  $[0, 1]^2$ , it exists  $N$ , large enough such that for every  $n \geq N$ ,  $\omega(x) \leq \varepsilon/4$  uniformly on  $[0, t]$ , thus  $\int_0^t \omega(x)(1 - G_0(x))dF_0(x) \leq \varepsilon/4$ . Therefore, for any  $t < \tau$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_0^t \omega(x)(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{2} \right) = 0, \quad (36)$$

For the second term of the right-hand side of Equation (35), note that  $2K(F_0(x), \widehat{F}(x-)) \leq K(F_0(x), F_0(x)) + K(\widehat{F}(x-), \widehat{F}(x-))$ , then  $\omega(x) \leq 2K(F_0(x), F_0(x)) + 2K(\widehat{F}(x-), \widehat{F}(x-))$ , and thus we deduce

$$\begin{aligned} & \mathbb{P} \left( \int_t^{\tau} \omega(x)(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{2} \right) \\ & \leq \mathbb{P} \left( \int_t^{\tau} K(F_0(x), F_0(x))(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{8} \right) + \mathbb{P} \left( \int_t^{\tau} K(\widehat{F}(x-), \widehat{F}(x-))(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{8} \right) \end{aligned} \quad (37)$$

By Condition 3 we deduce that it exists  $t' < \tau$ ,  $\int_{t'}^{\tau} K(F_0(x), F_0(x))(1 - G_0(x))dF_0(x) < \varepsilon/16$ . Then, by combining Equations (36) and (37), we get that for every  $t \geq t'$

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_0^{\tau} \omega(x)(1 - G_0(x))dF_0(x) \geq \varepsilon \right) \leq \limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_t^{\tau} K(\widehat{F}(x-), \widehat{F}(x-))(1 - G_0(x))dF_0(x) \geq \frac{\varepsilon}{8} \right),$$

and by taking  $t \rightarrow \tau$ , we deduce that  $\limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_0^{\tau} \omega(x)(1 - G_0(x))dF_0(x) \geq \varepsilon \right) = 0$  by Condition (4), completing our proof.  $\square$

**Part 2): Replacement of  $\frac{n}{n_0 n_1}L(x)$  by  $\psi(x)S_0(x)$**

By the previous part, our test statistic satisfies

$$nZ_n = n \left( \frac{n}{n_0 n_1} \right)^2 \int_0^{\tau_n} \int_0^{\tau_n} K(F_0(x), F_0(y))L(x)L(y)d\Lambda^*(x)d\Lambda^*(y) + o_p(1).$$

The next step is to show that we can replace  $n/(n_0 n_1)L(x)$  by its limit  $\psi(x)S_0(x)$  without altering the value of  $nZ_n$  by more than an error of order  $o_p(1)$ . We formalise this result in the following lemma.

**Lemma 20.** *Under Conditions 3 and 4, it holds*

$$nZ_n = n \int_0^{\tau_n} \int_0^{\tau_n} K(F(x), F(y))\psi(x)\psi(y)S_0(x)S_0(y)d\Lambda^*(x)d\Lambda^*(y) + o_p(1).$$

*Proof.* Similar to the proof of Lemma 19, we just need to prove that

$$n \left\| \int_0^{\tau_n} K(F(x), \cdot) \left( \frac{n}{n_0 n_1} L(x) - \psi(x) S_0(x) \right) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = o_p(1), \quad (38)$$

and

$$n \left\| \int_0^{\tau_n} K(F(x), \cdot) \psi(x) S_0(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 = O_p(1). \quad (39)$$

We begin by proving Equation (38). Note that, by expanding the inner product, it is enough to prove that

$$n \int_0^{\tau_n} \int_0^{\tau_n} h(x, y) d\Lambda^*(x) d\Lambda^*(y) = o_p(1),$$

where

$$h(x, y) = K(F(x), F(y)) \left( \frac{n}{n_0 n_1} L(x) - \psi(x) S_0(x) \right) \left( \frac{n}{n_0 n_1} L(y) - \psi(y) S_0(y) \right).$$

Also, observe that  $(h(x, y))_{x, y \geq 0}$  is a predictable positive-definite process. Then, by Theorem 17, we just need to verify that  $n \int_0^{\tau_n} h(x, y) d\Lambda^*(x) = o_p(1)$ , which is deduced from the following equalities:

$$\begin{aligned} n \int_0^{\tau_n} h(x, y) d\Lambda^*(x) &= n \int_0^{\tau_n} K(F(x), F(x)) \left( 1 - \frac{\psi(x) S_0(x)}{n/(n_0 n_1) L(x)} \right)^2 \left( \frac{n}{n_0 n_1} \right)^2 L(x) d\Lambda_0(x) \\ &= O_p(1) \int_0^{\tau} K(F(x), F(x)) \left( 1 - \frac{\psi(x) S_0(x)}{n/(n_0 n_1) L(x)} \right)^2 \psi(x) dF_0(x) \\ &= o_p(1) \end{aligned}$$

where the second equality is due to Propositions 11.ii and 11.iii, from which we deduce that  $L(x)/n = \Theta_p(1) \psi(x) S_0(x)$  uniformly for all  $x \leq \tau_n$ , and the third equality is due to an application of the dominated convergence theorem in sets of arbitrarily large probability: observe that, by Proposition 10, we have that  $\left( 1 - \frac{\psi(x) S_0(x)}{n/(n_0 n_1) L(x)} \right) \rightarrow 0$  for all  $x < \tau$ . Additionally,

$$\begin{aligned} K(F(x), F(x)) \left( 1 - \frac{\psi(x) S_0(x)}{n/(n_0 n_1) L(x)} \right)^2 \psi(x) &= O_p(1) K(F(x), F(x)) \psi(x) dF_0(x) \\ &= O_p(1) K(F(x), F(x)) (1 - G_0(x)), \end{aligned}$$

since  $\psi(x) \leq (1 - G_0(x))$  for all  $x < \tau$ , and note that  $K(F(x), F(x))(1 - G_0(x))$  is integrable by Condition 3. With these ingredients we can apply the dominated convergence theorem in set of arbitrarly large probability.

To check Equation (39), we follow the same steps, using that  $n \int_0^{\tau_n} K(F(x), F(x)) \psi(x)^2 S_0(x)^2 / L_0(x) d\Lambda_0(x) = O_p(1)$ , since  $L_0(x)/n = \Theta_p(1) \psi(x) S_0(x)$  uniformly for all  $x \leq \tau_n$ .  $\square$

**Part 3): Replacement of  $Y_0$  by  $\eta n(1 - H_0)$  and  $Y_1$  by  $(1 - \eta)n(1 - H_1)$**  From the previous step, it holds that

$$\begin{aligned} nZ_n &= n \left\| \int_0^{\tau_n} K(F_0(x), \cdot) \psi(x) S_0(x) d\Lambda^*(x) \right\|_{\mathcal{H}}^2 + o_p(1) \\ &= \frac{1}{n} \left\| \int_0^{\tau_n} K(F_0(x), \cdot) \psi(x) S_0(x) \left( \frac{dM_0(x)}{Y_0(x)/n} - \frac{dM_1(x)}{Y_1(x)/n} \right) \right\|_{\mathcal{H}}^2 + o_p(1) \end{aligned}$$

Our next step is to replace  $Y_0/n$  and  $Y_1/n$  by their corresponding limits  $\eta S_0(1 - G_0)$  and  $(1 - \eta)S_0(1 - G_1)$ . The following Lemma implies the desired result.

**Lemma 21.** *Under Conditions 3, it holds*

$$\frac{1}{n} \left\| \int_0^{\tau_n} K(F_0(x), \cdot) \psi(x) S_0(x) \left( \frac{1}{(Y_c(x)/n)} - \frac{\eta^{c-1} (1 - \eta)^{-c}}{S_0(x) (1 - G_c(x))} \right) dM_c(x) \right\|_{\mathcal{H}}^2 = o_p(1), \quad (40)$$

and

$$\frac{1}{n} \left\| \int_0^{\tau_n} K(F_0(x), \cdot) \frac{\psi(x)}{(1 - G_c(x))} dM_c(x) \right\|_{\mathcal{H}}^2 = O_p(1), \quad (41)$$

for any class label  $c \in \{0, 1\}$ .



*Proof.* We only prove the result for  $c = 0$  as the proof for  $c = 1$  follows the same steps. Define  $\sigma_n(x) = \left( \frac{1}{Y_0(x)/n} - \frac{\eta^{-1}}{S_0(x)(1-G_0(x))} \right)$ , and notice that

$$\begin{aligned} & \frac{1}{n} \left\| \int_0^{\tau_n} K(F_0(x), \cdot) \psi(x) S_0(x) \left( \frac{1}{(Y_0(x)/n)} - \frac{\eta^{-1}}{S_0(x)(1-G_0(x))} \right) dM_0(x) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n} \int_0^{\tau_n} \int_0^{\tau_n} K(F_0(x), F_0(y)) \psi(x) \psi(y) S_0(x) S_0(y) \sigma_n(x) \sigma_n(y) dM_0(x) dM_0(y) \end{aligned} \quad (42)$$

Define  $h(x, y)$  as the integrand in the previous double integral, and notice that  $(h(x, y))_{x, y \geq 0}$  is a predictable positive-definite process. Also, recall that  $M_0(x)$  is an  $(\mathcal{F}_x)_{x \geq 0}$ -martingale with predictable and quadratic variation processes given by  $d\langle M_0 \rangle(x) = Y_0(x) d\Lambda_0(x)$  and  $d[M_0](x) = dN(x)$ . Then, by Theorem 17, we just need to verify that

$$\frac{1}{n} \int_0^{\tau_n} h(x, y) d\langle M_0 \rangle(x) = o_p(1).$$

By Proposition 11.ii, we have  $Y_0(x)/n = O_p(1)S_0(x)(1-G_0(x))$  uniformly for all  $x \leq \tau_n$ .

$$\int_0^{\tau_n} h(x, y) d\langle M_0 \rangle(x) = O_p(1) \int_0^{\tau} K(F_0(x), F_0(x)) \psi(x)^2 \sigma_n(x)^2 \mathbb{1}_{\{x \leq \tau_n\}} S_0(x)^2 (1-G_0(x)) dF_0(x).$$

To prove that the previous expression is  $o_p(1)$  we use the dominated convergence theorem. For such, observe that  $\sigma_n(x) \rightarrow 0$  a.s. for all  $x < \tau$ , and that  $\sigma_n(x)^2 = O_p(1)(S_0(x)(1-G_0(x)))^2$  uniformly for all  $x \leq \tau_n$  due to proposition 11.iii, therefore

$$K(F_0(x), F_0(x)) \psi(x)^2 \sigma_n(x)^2 S_0(x)^2 (1-G_0(x)) = O_p(1) K(F_0(x), F_0(x)) \frac{\psi(x)^2}{1-G_0(x)},$$

uniformly for all  $x \leq \tau$ . Finally, note that  $K(F_0(x), F_0(x)) \frac{\psi(x)^2}{1-G_0(x)}$  is integrable by Condition 3, since  $\psi(x) \leq \min\{\eta^{-1}(1-G_1(x)), (1-\eta)^{-1}(1-G_0(x))\}$ . The previous analysis yields that we can use the dominated convergence theorem in sets of large probability, concluding that Equation (40) holds true. Equation (41) follows from the same arguments, by noticing that  $\int_0^{\tau_n} K(F_0(x), F_0(x)) \frac{\psi(x)}{(1-G_0(x))} Y_0(x) d\Lambda_0(x) = O_p(1)$ .  $\square$

### B.3 Step II: Asymptotically equivalence of two models

After applying Theorem 18 we get

$$Z_n = \frac{1}{n^2} \sum_i^n \sum_{j=1}^n \int_0^{\tau_n} \int_0^{\tau_n} \frac{K(F_0(x), F_0(y)) \psi(x) \psi(y)}{(1-G_{c_i}(x))(1-G_{c_j}(y))} \eta^{-2} \left( \frac{-\eta}{(1-\eta)} \right)^{c_i+c_j} dM^i(x) dM^j(y) + o_p(n^{-1}), \quad (43)$$

Recall that the previous expression is valid when considering either deterministic or independent Bernoulli( $\eta$ ) covariates. In this section we show that asymptotic results obtained using either the deterministic or random covariates models are equivalent.

Let  $s_i = (X_i, \Delta_i, c_i)$  for all  $i \in [n]$  and define  $J : (\mathbb{R}_+ \times \{0, 1\} \times \{0, 1\})^2 \rightarrow \mathbb{R}$  as

$$J(s_i, s_j) = \eta^{-2} \left( \frac{-\eta}{(1-\eta)} \right)^{c_i+c_j} \int_0^{X_i} \int_0^{X_j} \frac{K(F_0(x), F_0(y)) \psi(x) \psi(y)}{(1-G_{c_i}(x))(1-G_{c_j}(y))} dM^i(x) dM^j(y). \quad (44)$$

Also, define

$$V_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n J(s_i, s_j), \quad (45)$$

and notice that  $V_n$  is the non-negligible part of  $Z_n$ , that is,  $Z_n = V_n + o_p(n^{-1})$  for both, the deterministic and the random covariates models. In this section only, we denote by  $Z'_n$  and  $Z_n$ , respectively, the test statistics under the deterministic and random covariates models. More generally, we use an apostrophe (e.g.,  $V'_n$ ) to denote any term related to the deterministic covariates model.

**Lemma 22.** *There exists a coupling of  $Z_n$  and  $Z'_n$  such that  $Z_n - Z'_n = o_p(n^{-1})$ .*

*Proof.* We construct a coupling such that  $V_n - V'_n = o(n^{-1})$ , which implies the desired result. The coupling is constructed as follows:

1. For  $c \in \{0, 1\}$ , denote by  $\mu_c$  the measure induced by the random pair  $(X, \Delta)$  where  $X = \min(T, C)$  and  $\Delta = \mathbb{1}_{\{T \leq C\}}$ , and  $T$  and  $C$  are independent random variables such that  $T \sim F_c$  and  $C \sim G_c$ .

2. Let  $\mathcal{L}_0 = (X_i, \Delta_i, 0)_{i \geq 1}$  and  $\mathcal{L}_1 = (\bar{X}_i, \bar{\Delta}_i, 1)_{i \geq 1}$  be such that all the triples in  $\mathcal{L}_0$  and  $\mathcal{L}_1$  are independent of everything, and  $(X_i, \Delta_i) \sim \mu_0$  and  $(\bar{X}_i, \bar{\Delta}_i) \sim \mu_1$  for all  $i \geq 1$ . Note that, since the times are sampled from continuous distributions, the data points are unique almost surely.
3. In the deterministic covariates model we generate the data, namely  $\mathcal{D}'_n$ , by choosing the first  $n_0$  elements of  $\mathcal{L}_0$  and the first  $n_1$  elements of  $\mathcal{L}_1$ .
4. In the random covariates model, we generate the data, namely  $\mathcal{D}_n$ , by sampling  $N_0 \sim \text{Binomial}(n, \eta)$  and then choosing the first  $N_0$  elements of  $\mathcal{L}_0$  and the first  $n - N_0$  elements of  $\mathcal{L}_1$ .
5. Compute  $V'_n$  and  $V_n$  by using the datasets  $\mathcal{D}'_n$  and  $\mathcal{D}_n$ , respectively.

Note that the datasets  $\mathcal{D}_n$  and  $\mathcal{D}'_n$  differ in exactly  $Q = |N_0 - n_0|$  points. For ease of notation, denote by  $s_i = (X_i, \Delta_i, c_i)$  the observations in  $\mathcal{D}_n$  and by  $s'_i = (X'_i, \Delta'_i, c'_i)$  the observations in  $\mathcal{D}'_n$ . We sort the datasets in such a way that the first  $Q$  elements of each dataset are not present in the other, and the remaining  $n - Q$  elements in such a way that  $s_j = s'_j$  for  $j \geq Q + 1$ . Then, a simple computation shows that

$$V_n - V'_n = D_n + R_n, \quad (46)$$

where

$$D_n = \frac{1}{n^2} \sum_{i=1}^Q \sum_{j=1}^Q J(s_i, s_j) - J(s'_i, s'_j) \quad \text{and} \quad R_n = \frac{2}{n^2} \sum_{i=1}^Q \sum_{j=Q+1}^n J(s_i, s_j) - J(s'_i, s'_j). \quad (47)$$

We continue by proving that  $nD_n = o_p(1)$  and  $nR_n = o_p(1)$ . To this end, we need the following intermediate result, whose proof is deferred to Appendix E.

**Proposition 23.** *Let  $A : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$  be a symmetric function such that, for every  $c, c' \in \{0, 1\}$ , the following integrals are finite:  $\int_0^\tau |A(x, x)|(1 - G_c(x))dF_c(x)$  and  $\int_0^\tau \int_0^\tau A(x, y)^2(1 - G_c(x))(1 - G_{c'}(y))dF_c(x)dF_{c'}(y)$ . Define  $J(s_i, s_j)$  by*

$$J(s_i, s_j) = \int_0^{X_i} \int_0^{X_j} A(x, y)dM^i(x)dM^j(y)$$

for all  $i, j \in [n]$ , and denote by  $\tilde{\mathbb{E}}$  the expectation conditioned on fixed covariate values  $c_i$  and  $c_j$ . Then the following hold:

- i)  $\tilde{\mathbb{E}}(J(s_i, s_i)) = \int_0^\tau A(x, x)(1 - G_{c_i}(x))dF_{c_i}(x)$ ,
- ii)  $\tilde{\mathbb{E}}(|J(s_i, s_i)|) < \infty$ ,
- iii) for all  $i \neq j$ ,  $\tilde{\mathbb{E}}(J(s_i, s_j)|s_j) = 0$ , and
- iv) for all  $i \neq j$ ,  $\tilde{\mathbb{E}}(J(s_i, s_j)^2) = \int_0^\tau \int_0^\tau A(x, y)^2(1 - G_{c_i}(x))(1 - G_{c_j}(y))dF_{c_i}(x)dF_{c_j}(y)$ .

In our applications of Proposition 23, we will choose

$$A(x, y) = \frac{1}{\eta^2} \left( \frac{-\eta}{1 - \eta} \right)^{c_k + c_\ell} \frac{K(F_0(x), F_0(y))\psi(x)\psi(y)}{(1 - G_k(x))(1 - G_\ell(y))}.$$

We continue by proving  $n|D_n| = o_p(1)$ . Observe that, by symmetry,

$$\begin{aligned} n|D_n| &= \frac{1}{n} \left| \sum_{i=1}^Q \sum_{j=1}^Q J(s_i, s_j) - J(s'_i, s'_j) \right| \\ &\leq \frac{1}{n} \left| \sum_{i=1}^Q J(s_i, s_i) - J(s'_i, s'_i) \right| + \frac{2}{n} \left| \sum_{i=1}^Q \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \right|. \end{aligned} \quad (48)$$

For the first term in Equation (48), notice that by conditioning on  $Q$ , we obtain

$$\begin{aligned} \mathbb{E} \left( \frac{1}{n} \left| \sum_{i=1}^Q J(s_i, s_i) - J(s'_i, s'_i) \right| \middle| Q \right) &\leq \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^Q |J(s_i, s_i)| \middle| Q \right) + \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^Q |J(s'_i, s'_i)| \middle| Q \right) \\ &= \frac{Q}{n} \mathbb{E}(|J(s_1, s_1)| | c_1 = 0) + \frac{Q}{n} \mathbb{E}(|J(s'_1, s'_1)| | c'_1 = 1), \end{aligned}$$

where the last equality holds since, without loss of generality, the first  $Q$  elements of  $\mathcal{D}_n$  are chosen from  $\mathcal{L}_0$  and the  $Q$  elements of  $\mathcal{D}'_n$  are chosen from  $\mathcal{L}_1$ . Additionally, notice that by Proposition 23.ii,  $\mathbb{E}(|J(s_1, s_1)| | c_1 = 0) < \infty$  and  $\mathbb{E}(|J(s'_1, s'_1)| | c'_1 = 1) < \infty$ .

Also, note that

$$\mathbb{E} \left( \frac{Q}{n} \right)^2 \leq \mathbb{E} \left( \frac{Q^2}{n^2} \right) \leq 2\mathbb{E} \left( \left| \frac{N_0}{n} - \eta \right|^2 \right) + 2 \left| \eta - \frac{n_0}{n} \right|^2 = 2\text{Var} \left( \frac{N_0}{n^2} \right) + 2 \left| \eta - \frac{n_0}{n} \right|^2 \rightarrow 0 \quad (49)$$

where the last limit holds because  $\text{Var}(N_0) = \eta(1-\eta)n$ , and  $n_0/n \rightarrow \eta$ . Then, by using the previous result, we deduce

$$\mathbb{E} \left( \frac{1}{n} \left| \sum_{i=1}^Q J(s_i, s_i) - J(s'_i, s'_i) \right| \right) = \mathbb{E} \left( \mathbb{E} \left( \frac{1}{n} \left| \sum_{i=1}^Q J(s_i, s_i) - J(s'_i, s'_i) \right| \middle| Q \right) \right) = O(1) \mathbb{E} \left( \frac{Q}{n} \right) \rightarrow 0.$$

We continue by proving that the second term in Equation (48) is  $o_p(1)$ . Let  $\varepsilon > 0$ , then

$$\begin{aligned} \mathbb{P} \left( \frac{2}{n} \left| \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \right| \geq \varepsilon \right) &= \mathbb{E} \left( \mathbb{P} \left( \frac{2}{n} \left| \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \right| \geq \varepsilon \middle| Q \right) \right) \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \middle| Q \right) \right), \end{aligned}$$

and

$$\begin{aligned} &\text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \middle| Q \right) \\ &= \text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) \middle| Q \right) + \text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) \middle| Q \right) \\ &= \frac{4}{n^2} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q \text{Var}(J(s_i, s_j)) + \text{Var}(J(s'_i, s'_j)) \\ &= \frac{4}{n^2} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q \mathbb{E}(J(s_i, s_j)^2) + \mathbb{E}(J(s'_i, s'_j)^2), \end{aligned}$$

the first inequality holds because the variables  $(s_i)_{i=1}^Q$  are independent of  $(s'_i)_{i=1}^Q$ , given  $Q$ . The second equality follows from noticing that  $\text{Cov}(J(s_i, s_j), J(s_l, s_k)) = 0$  for all  $(i, j) \neq (l, k)$ , which is a simple computation that follows straightforwardly from Proposition 23.iii. By Proposition 23.iv, there exists  $C' > 0$  such that for all  $i \neq j$ ,  $\mathbb{E}(J(s_i, s_j)^2) \leq C'$  and  $\mathbb{E}(J(s'_i, s'_j)^2) \leq C'$ , then

$$\text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \middle| Q \right) \leq C \frac{Q^2}{n^2},$$

for some  $C > 0$ . We conclude

$$\begin{aligned} \mathbb{P} \left( \frac{2}{n} \left| \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \right| \geq \varepsilon \right) &\leq \frac{1}{\varepsilon^2} \mathbb{E} \left( \text{Var} \left( \frac{2}{n} \sum_{i=1}^{Q-1} \sum_{j=i+1}^Q J(s_i, s_j) - J(s'_i, s'_j) \middle| Q \right) \right) \\ &\leq \frac{C}{\varepsilon^2} \mathbb{E}(Q^2/n^2) \rightarrow 0, \end{aligned}$$

where the last limit follows from Equation (49).

Finally, we prove  $n|R_n| = o_p(1)$ . Let  $\varepsilon > 0$ , then

$$\mathbb{P}(n|R_n| \geq \varepsilon) = \mathbb{E}(\mathbb{P}(n|R_n| \geq \varepsilon | Q)) \leq \frac{1}{\varepsilon^2} \mathbb{E}(\text{Var}(nR_n | Q))$$

and

$$\text{Var}(nR_n | Q) = \frac{4}{n^2} \sum_{i=1}^Q \sum_{j=Q+1}^n \text{Var}(J(s_i, s_j) | Q) + \text{Var}(J(s'_i, s'_j) | Q) \leq C \frac{Q(n-Q)}{n^2},$$

where the previous equality follows by Proposition 23.iii, from which we deduce that all non-trivial covariances are 0, and the inequality holds for some constant  $C > 0$  due to Proposition 23.iv, as indeed,  $\text{Var}(J(s_i, s_j) | Q)$  and  $\text{Var}(J(s'_i, s'_j) | Q)$  are bounded for any  $i, j \in [n]$ , independently of the value of the covariates. Since  $\mathbb{E}(Q)/n \rightarrow 0$  we conclude that  $nR_n = o_p(1)$  which, combined with the fact that  $nD_n = o_p(1)$ , yields  $n(V_n - V'_n) = o_p(1)$ .  $\square$

## B.4 Step III: Limit distribution under the null hypothesis

In this section we prove Theorem 5 for the random covariates model. Recall that by Lemma 22, the result also holds for the deterministic covariates model.

Recall from Equation (45) that

$$\frac{n_0 n_1}{n} Z_n = \frac{n_0 n_1}{n} V_n + o_p(1) = \frac{n_0 n_1}{n} \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n J(s_i, s_j) + o_p(1), \quad (50)$$

where  $J$  is defined in Equation (44) and  $s_i = (X_i, \Delta_i, c_i)$  are independent and identically distributed random observations as the covariates are random, which is the main advantage of working in the random covariates model. Then, by using the previous expression, the asymptotic distribution of  $(n_0 n_1)/n Z_n$  can be obtained from a standard application of the theory of V-statistics. In particular, by Proposition 23.iii,  $\mathbb{E}(J(s_i, s_j)|s_j) = 0$  for any  $i \neq j$ , hence  $J$  is a so-called *degenerate* V-statistic kernel, therefore Theorem 4.3.2 of Koroljuk and Borovskich [1994] yields

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n J(s_i, s_j) \xrightarrow{D} \mathbb{E}(J(s_1, s_1)) + \mathcal{Y}, \quad (51)$$

where  $\mathcal{Y} = \sum_{i=1}^{\infty} \lambda_i (\xi_i^2 - 1)$ ,  $(\xi_i)_{i \geq 1}$  are independent and identically distributed standard normal random variables, and  $(\lambda_i)_{i \geq 1}$  are the eigenvalues associated to the integral operator  $T_J : L^2(X_1, \Delta_1, c_1) \rightarrow L^2(X_1, \Delta_1, c_1)$  defined by  $(T_J \phi)(s) = \mathbb{E}_{s_1}(J(s, s_1)\phi(s_1))$ , where  $\mathbb{E}_{s_1}$  means expectation with respect  $s_1 = (X_1, \Delta_1, c_1)$ . From Equation (51) we deduce that

$$\frac{n_0 n_1}{n} Z_n \xrightarrow{D} \eta(1 - \eta)\mathbb{E}(J(s_1, s_1)) + \eta(1 - \eta)\mathcal{Y}.$$

We finish by computing the asymptotic mean and variance of our test statistic. By using Proposition 23.i, a simple computation shows that

$$\mathbb{E}(J(s_1, s_1)) = \frac{1}{\eta} \int_0^\tau \frac{K(F_0(x), F_0(x))\psi(x)^2}{1 - G_0(x)} dF_0(x) + \frac{1}{1 - \eta} \int_0^\tau \frac{K(F_0(x), F_0(x))\psi(x)^2}{1 - G_1(x)} dF_0(x),$$

and since  $\mathcal{Y}$  has mean 0, the asymptotic mean of  $(n_0 n_1)/n Z_n$  is given by

$$\eta(1 - \eta)\mathbb{E}(J(s_1, s_1)) = \int_0^\tau K(F_0(x), F_0(x))\psi(x) dF_0(x).$$

Finally, from Proposition 23.iv, we have that

$$\text{Var}(\mathcal{Y}) = 2 \sum_{i=1}^{\infty} \lambda_i^2 = 2\mathbb{E}(J(s_1, s_2)^2) = \frac{2}{(\eta(1 - \eta))^2} \int_0^\tau \int_0^\tau K(F_0(x), F_0(y))^2 \psi(x)\psi(y) dF_0(x) dF_0(y),$$

from which we deduce that the expression for the asymptotic variance is given by

$$\text{Var}(\eta(1 - \eta)\mathcal{Y}) = 2 \int_0^\tau \int_0^\tau K(F_0(x), F_0(y))^2 \psi(x)\psi(y) dF_0(x) dF_0(y).$$

## C Wild Bootstrap

By following exactly the same steps used in the proof of Theorem 5, the Wild Bootstrap test statistic  $Z_n^{\mathcal{W}}$ , given in Equation (21), can be rewritten as  $Z_n^{\mathcal{W}} = V_n^{\mathcal{W}} + o_p(n^{-1})$ , where  $V_n^{\mathcal{W}} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n W_i W_j J(s_i, s_j)$ , and  $J$  is the kernel defined in Equation (44). Under the random covariates model  $V_n$  is a degenerate V-statistic, and thus, Theorem 9 is a direct application of Theorem 3.1 of Dehling and Mikosch [1994]. This result can be extended to the deterministic covariates model by using the same coupling of Lemma 22 (however, randomness is taken over  $\mathcal{W}$ , as we are conditioning on the sequence of data points).

## D Proof of Theorem 6 and Corollary 7

In this section we mainly work under the alternative hypothesis, therefore we need to consider this fact when computing some limit results. For instance, in this more general setting the limit of  $n/(n_0 n_1)L(x)$  is slightly different to its limit under the null hypothesis. Indeed, by Proposition 10, for every fixed  $x$ , it holds

$$\frac{n}{n_0 n_1} L(x) \rightarrow \psi^*(x) S_1(x) S_0(x) \quad a.s., \quad (52)$$

as the number of observations tends to infinity, and by Proposition 11

$$\frac{n}{n_0 n_1} L(x) = O_p(1) \psi^*(x) S_1(x) S_0(x), \quad (53)$$

uniformly for all  $x \leq \tau_n$ . In the previous results,  $\psi^*(x)$  is the function defined in Equation (15) which satisfies the following bound

$$\psi^*(x) \leq \min \left\{ \frac{1 - G_0(x)}{\eta S_1(x)}, \frac{1 - G_1(x)}{(1 - \eta) S_0(x)} \right\}. \quad (54)$$

Other objects that may vary their limits in this more general setting are the pooled Nelson-Aalen estimator  $\widehat{\Lambda}$  and Kaplan-Meier statistic  $\widehat{F}$ , and in general, any statistic constructed using the survival times  $(T_i)_{i \in [n]}$  (as now,  $F_0$  and  $F_1$  may or may not be equal). Having this in mind, we proceed to prove Theorem 6 and Corollary 7.

To proof Theorem 6, we require the following intermediate result:

**Lemma 24.** *Assume that Conditions 3 and 4 hold. Then,  $\|\phi_c^n - \phi_c\|_{\mathcal{H}} \xrightarrow{\mathbb{P}} 0$  for  $c \in \{0, 1\}$ .*

*Proof.* We prove the result for  $c = 0$ , as the proof for  $c = 1$  is exactly the same. We consider the deterministic covariates model since for the random covariates model we can condition on the number of data points with covariate equal to 0.

Define the intermediate embeddings,  $\widehat{\phi}_0 \in \mathcal{H}$  and  $\phi'_0 \in \mathcal{H}$ , by

$$\widehat{\phi}_0(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(y-), \cdot) L(y) d\Lambda_0(y), \quad \text{and} \quad \phi'_0(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(F(y), \cdot) L(y) d\Lambda_0(y),$$

and recall that  $\phi_0^n$  and  $\phi_0$  (introduced in Equations (10) and (18), respectively) are given by

$$\phi_0^n(\cdot) = \frac{n}{n_0 n_1} \int_0^{\tau_n} K(\widehat{F}(y-), \cdot) L(y) d\widehat{\Lambda}_0(y) \quad \text{and} \quad \phi_0(\cdot) = \int_0^{\tau} K(F(y), \cdot) \psi^*(y) S_1(y) dF_0(y).$$

Notice that in this setting the pooled distribution  $F$  is not necessarily equal to  $F_0$ .

Clearly

$$\|\phi_0^n - \phi_0\|_{\mathcal{H}} \leq \|\phi_0^n - \widehat{\phi}_0\|_{\mathcal{H}} + \|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}} + \|\phi'_0 - \phi_0\|_{\mathcal{H}}.$$

We will prove that the three terms in the right-hand-side of the previous equation are  $o_p(1)$ . First we prove  $\|\phi_0^n - \widehat{\phi}_0\|_{\mathcal{H}} = o_p(1)$ . Recall that the Nelson-Aalen estimator satisfies  $(d\widehat{\Lambda}_0 - d\Lambda_0)(x) = dM_0(x)/Y_0(x)$ , then

$$\|\phi_0^n - \widehat{\phi}_0\|_{\mathcal{H}}^2 = \left( \frac{n}{n_0 n_1} \right)^2 \left\| \int_0^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) \frac{dM_0(x)}{Y_0(x)} \right\|_{\mathcal{H}}^2 = \frac{1}{n_0^2} \int_0^{\tau_n} \int_0^{\tau_n} h(x, y) dM_0(x) dM_0(y), \quad (55)$$

where

$$h(x, y) = K(\widehat{F}(x-), \widehat{F}(y-)) \frac{(Y_1(x)/n_1)(Y_1(y)/n_1)}{(Y(x)/n)(Y(y)/n)}.$$

We prove that Equation (55) is  $o_p(1)$  by using Theorem 17. Observe that  $h(x, y)$  is symmetric and positive-definite,  $(h(x, y))_{(x, y) \in C}$  is  $\mathcal{P}$  measurable by Proposition 14, and  $(h(x, x))_{x \geq 0}$  is predictable with respect to  $(\mathcal{F}_x)_{x \geq 0}$ . Then, by Theorem 17, the result is deduced from checking that  $\frac{1}{n_0^2} \int_0^{\tau_n} h(x, x) d\langle M_0 \rangle(x) = o_p(1)$ . By using that  $Y_1(x)/Y(x) \leq 1$ ,  $n/n_1 = O(1)$ , and that  $Y_0(x)/n_0 = O_p(1)S(x)(1 - G_0(x))$  uniformly for all  $x \leq \tau_n$ , we have

$$\begin{aligned} \frac{1}{n_0^2} \int_0^{\tau_n} h(x, x) d\langle M_0 \rangle(x) &= \frac{1}{n_0^2} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) \frac{(Y_1(x)/n_1)^2}{(Y(x)/n)^2} Y_0(x) d\Lambda_0(x) \\ &= O(1) \frac{1}{n_0} \int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) (1 - G_0(x)) dF_0(x). \end{aligned}$$

The previous term tends to 0 since  $\int_0^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) (1 - G_0(x)) dF_0(x) < \infty$  by Condition 4.

We continue by proving  $\|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}} = o_p(1)$ . Let  $t < \tau$  and  $\epsilon > 0$ , then

$$\begin{aligned} &\mathbb{P} \left( \|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}}^2 > \epsilon \right) \\ &\leq \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_0^t h(x, \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}}^2 > \epsilon/2 \right) + \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} h(x, \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}}^2 > \epsilon/2 \right) \end{aligned}$$

where  $h(x, \cdot) = (K(\widehat{F}(x-), \cdot) - K(F(x), \cdot))$ . By taking limsup when  $n$  grows to infinity, the first probability is equal to zero since  $h(x, y) = \langle h(x, \cdot), h(y, \cdot) \rangle_{\mathcal{H}} \rightarrow 0$  uniformly for all  $x, y \leq t$ . Then,

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}}^2 > \epsilon \right) \leq \limsup_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} h(x, \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}}^2 > \epsilon/2 \right).$$

Finally, by taking limit when  $t \rightarrow \tau$ , and by using the definition of  $h$ , we obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}}^2 > \epsilon \right) &\leq \lim_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}} > \sqrt{\epsilon/2}/2 \right) \\ &+ \lim_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(F(x), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}} > \sqrt{\epsilon/2}/2 \right). \end{aligned}$$

We prove that each of the terms in the right-hand-side of the previous Equation are 0. For the first one, from Propositions 11.ii and 11.iii, we have that  $(n/(n_0 n_1))L(x) = \beta^{-1} \psi^*(x) S_0(x) S_1(x)$  for all  $x \leq \tau_n$  with probability at least  $1 - p(\beta)$  with  $p(\beta) \rightarrow 0$  as  $\beta \rightarrow 0$ . Then, with probability at least  $1 - p(\beta)$  we have

$$\begin{aligned} \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(F(x), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}}^2 &\leq \beta^{-2} \int_t^{\tau_n} \int_t^{\tau_n} K(\widehat{F}(x-), \widehat{F}(y-)) S_1(x) \psi^*(x) S_1(y) \psi^*(y) dF_0(x) dF_0(y) \\ &\leq 2\beta^{-2} \int_t^{\tau_n} \int_t^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) S_1(x) \psi^*(x) S_1(y) \psi^*(y) dF_0(x) dF_0(y) \\ &\leq 2\beta^{-2} \int_t^{\tau_n} S_1(y) \psi^*(y) dF_0(y) \int_t^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) S_1(x) \psi^*(x) dF_0(x) \\ &\leq \frac{2}{\eta^2} \beta^{-2} \int_t^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) dF_0(x), \end{aligned}$$

where the second inequality holds by noting that  $2K(s, t) \leq K(s, s) + K(t, t)$  for all  $s, t$  since  $K$  is positive definite, and the fourth inequality holds Equation (54). Finally,

$$\begin{aligned} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}} > \sqrt{\epsilon/2}/2 \right) \\ \leq p(\beta) + \mathbb{P} \left( \int_t^{\tau_n} K(\widehat{F}(x-), \widehat{F}(x-)) dF_0(x) \geq \beta^2 \eta^2 \sqrt{\epsilon/2}/4 \right) \end{aligned} \quad (56)$$

then, by Condition 4, we conclude that for all  $\beta \in (0, 1)$

$$\limsup_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(\widehat{F}(x-), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}} > \sqrt{\epsilon/2}/2 \right) \leq p(\beta), \quad (57)$$

then taking  $\beta \rightarrow 0$  yields the desired result. A similar argument shows that

$$\lim_{t \rightarrow \tau} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \left\| \frac{n}{n_0 n_1} \int_t^{\tau_n} K(F(x), \cdot) L(x) d\Lambda_0(x) \right\|_{\mathcal{H}} > \sqrt{\epsilon/2}/2 \right) = 0,$$

concluding that  $\|\widehat{\phi}_0 - \phi'_0\|_{\mathcal{H}}^2 = o_p(1)$ .

Finally, we prove that  $\|\phi'_0 - \phi_0\|_{\mathcal{H}} = o_p(1)$ . Observe that by triangular inequality,

$$\begin{aligned} \|\phi'_0 - \phi_0\|_{\mathcal{H}} \\ \leq \left\| \int_0^{\tau_n} K(x, \cdot) \left( \frac{n}{n_0 n_1} L(x) - \psi^*(x) S_1(x) S_0(x) \right) d\Lambda_0(x) \right\|_{\mathcal{H}} + \left\| \int_{\tau_n}^{\tau} K(F(y), \cdot) \psi^*(y) S_1(y) dF_0(y) \right\|_{\mathcal{H}}. \end{aligned}$$

Observe that

$$\begin{aligned} \left\| \int_0^{\tau_n} K(x, \cdot) \left( \frac{n}{n_0 n_1} L(x) - \psi^*(x) S_1(x) S_0(x) \right) d\Lambda_0(x) \right\|_{\mathcal{H}}^2 &= \int_0^{\tau_n} \int_0^{\tau_n} K(x, y) \sigma(x) \sigma(y) \frac{dF_0(x)}{S_0(x)} \frac{dF_0(y)}{S_0(y)} \\ &= o_p(1) \end{aligned}$$

where  $\sigma(x) = \psi^*(x) S_0(x) S_1(x) - n/(n_0 n_1) L(x)$ . The second equality follows from the dominated convergence theorem: by Equation (52), we have  $\sigma(x) \rightarrow 0$  for all  $x \leq \tau$  almost surely, also, by Equations (53) and (54), it holds  $\sigma(x)/S_0(x) = O_p(1) \psi^*(x) S_1(x) = O_p(1)(1 - G_0(x))$  uniformly for all  $x \leq \tau_n$ . From Condition 3 we have that  $\int_0^{\tau} \int_0^{\tau} |K(F(x), F(y))| (1 - G_0(x))(1 - G_0(y)) dF_0(x) dF_0(y) < \infty$ , then we can apply the dominated convergence theorem on sets of probability as high as desired, obtaining that  $\|\phi'_0 - \phi_0\|_{\mathcal{H}} = o_p(1)$ .

Finally, observe that

$$\begin{aligned} \left\| \int_{\tau_n}^{\tau} K(F(y), \cdot) \psi^*(y) S_1(y) dF_0(y) \right\|_{\mathcal{H}}^2 &= \int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} K(F(x), F(y)) \psi^*(x) \psi^*(y) S_1(x) S_1(y) dF_0(x) dF_0(y) \\ &= O(1) \int_{\tau_n}^{\tau} \int_{\tau_n}^{\tau} K(F(x), F(y)) (1 - G(x)) (1 - G(y)) dF_0(x) dF_0(y) \\ &= o_p(1) \end{aligned}$$

where the second equality is due to Equation (54), and the third is due Condition 3 together with the fact that  $\tau_n \rightarrow \tau$ .  $\square$

*Proof of Theorem 6.* Notice that  $\sqrt{Z_n} = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}$ . Then, by using the triangle inequality, we deduce

$$\|\phi_0 - \phi_1\| - \|\phi_0 - \phi_0^n\|_{\mathcal{H}} - \|\phi_1 - \phi_1^n\|_{\mathcal{H}} \leq \sqrt{Z_n} \leq \|\phi_0^n - \phi_0\|_{\mathcal{H}} + \|\phi_0 - \phi_1\|_{\mathcal{H}} + \|\phi_1 - \phi_1^n\|_{\mathcal{H}},$$

hence, a straightforward application of Lemma 24 deduces  $\sqrt{Z_n} \xrightarrow{\mathbb{P}} \|\phi_0 - \phi_1\|_{\mathcal{H}}$ , and by squaring both sides, we obtain the desired result.  $\square$

*Proof of Corollary 7.* Theorem 6 yields  $\sqrt{Z_n} \xrightarrow{\mathbb{P}} \|\phi_0 - \phi_1\|_{\mathcal{H}}$ , therefore it is enough to prove that  $\phi_0 \neq \phi_1$ . Recall from Equation (18) that for  $c \in \{0, 1\}$ ,  $\phi_c$  is defined by

$$\phi_c(\cdot) = \int_0^{\tau} K(F(y), \cdot) d\nu_c(y)$$

where the measures  $\nu_c$  is defined in Equation (17). Let  $F^{-1}(x) = \inf\{t \geq 0 : F(t) = x\}$ , which is differentiable almost everywhere since  $F$  is the cumulative distribution function of a continuous random variable. Then, by performing a change of variables, we obtain

$$\phi_c(\cdot) = \int_0^1 K(x, \cdot) \frac{1}{f(F^{-1}(x))} d\nu_c(F^{-1}(x)),$$

where  $f$  is the density function associated with  $F$ . We conclude that  $\phi_c$  is the mean kernel embedding (see Equation (6)) of the measure  $\pi_c(\cdot)$  on  $[0, 1]$  given by  $\pi_c(A) = \int_A \frac{1}{f(F^{-1}(x))} d\nu_c(F^{-1}(x))$ .

As the kernel  $K$  is continuous and c-universal (recall the definition from Section 3.1), the mean kernel embedding of finite signed measures is injective, hence, we just need to check that  $\pi_0 \neq \pi_1$  and that both measures are finite.

First, we prove that  $F_0 \neq F_1$  implies that  $\pi_0 \neq \pi_1$ . We proceed by contradiction, i.e., we assume that  $\pi_0 = \pi_1$ . Since  $F$  and  $F^{-1}$  are non-decreasing functions,  $\pi_0 = \pi_1$  implies  $\nu_0 = \nu_1$ . Observe that  $\nu_0$  and  $\nu_1$  are continuous measures with respective densities  $\psi^*(x)S_1(x)f_0(x)$  and  $\psi^*(x)S_0(x)f_1(x)$ , where  $f_0$  and  $f_1$  denote the density functions of  $F_0$  and  $F_1$ , respectively. Since  $\nu_0 = \nu_1$ , their densities are equal. Let  $\tau_0 = \sup\{t : S_0(t) > 0\}$  and  $\tau_1 = \sup\{t : S_1(t) > 0\}$ , and, without loss of generality, assume that  $\tau_0 \leq \tau_1$ . By hypothesis, for all  $x < \tau_0$  we have that  $G_0(x) < 1$  and  $G_1(x) < 1$ , hence  $\psi^*(x) > 0$ . Using that  $\psi^*(x)S_1(x)f_0(x) = \psi^*(x)S_0(x)f_1(x)$ , we deduce that for all  $x < \tau_0$ ,  $S_1(x)f_0(x) = S_0(x)f_1(x)$ ; thus,  $\lambda_0(x) = \lambda_1(x)$ , and therefore, by the bijection between hazard and density functions, we have  $F_0(x) = F_1(x)$ . Since  $F_0(\tau_0) = 1$ , we conclude that  $F_0(x) = F_1(x)$  for all  $x \in \mathbb{R}_+$ , which is a contradiction.

Finally, a simple computation verifies that  $\nu_c$  is a finite measure. Indeed, for  $c = 0$ , we have  $\int_0^{\tau} d\nu_0(x) = \int_0^{\tau} \psi^*(x)S_1(x)f_0(x)dx$ , and from Equation (54) it holds that  $\psi^*(x)S_1(x)f_0(x) \leq C f_0$ , where  $C > 0$ , therefore  $\nu_0$  is a finite measure. The same holds for  $\nu_1$ , concluding the result.  $\square$

## E Deferred proofs

### E.1 Proof of Theorem 17

To ease notation, we write  $h$  and  $W$  instead of  $h_n$  and  $W_n$ , respectively. Define the process  $R$  by

$$R(t) = \int_0^t \int_0^t h(x, y) dW(x) dW(y), \quad t \geq 0. \quad (58)$$

Observe that  $R(0) = 0$  and that  $R(t) \geq 0$  for every  $t \geq 0$  since  $h$  is positive definite. We will prove that exists an increasing predictable process  $A$  that majorises  $R$  in the sense that  $\mathbb{E}(R(T)) \leq \mathbb{E}(A(T))$  for every finite stopping time, and then, we will apply the Lenglart-Rebolledo inequality to  $R$  and  $A$ , which tells us that for every  $\varepsilon, \delta > 0$ ,  $\mathbb{P}(\sup_{t \leq T} R(t) \geq \varepsilon) \leq \varepsilon/\delta + \mathbb{P}(A(T) \geq \delta)$  for any stopping time  $T$  (even unbounded).

To find the process  $A$ , we start by noticing that  $R$  can be decomposed as  $R(t) = D(t) + 2Z(t)$ , where the processes  $D$  and  $Z$  are given by

$$D(t) = \int_0^t h(x, x)d[W](x), \text{ and } Z(t) = \int_0^t \int_{(0, y)} h(x, y)dW(x)dW(y),$$

respectively. We make two observations. First, since the process  $(h(x, y))_{(x, y) \in C}$  is  $\mathcal{P}$ -measurable, then  $(Z(t))_{t \geq 0}$  is an  $(\mathcal{F}_t)$ -martingale by Theorem 15. Additionally, by the Optional Stopping Theorem,  $Z(T) = 0$  for every bounded stopping time  $T$ . Second, since  $h$  is positive definite,  $h(t, t) \geq 0$  and  $(h(x, x))_{x \geq 0}$  is  $(\mathcal{F}_x)$ -predictable, then the process  $(D(t))_{t \geq 0}$  is increasing and adapted with  $D(0) = 0$  and it is compensated by the process  $A$  given by

$$A(t) = \int_0^t h(Q(x), Q(x))d\langle W \rangle \quad (59)$$

which is increasing and predictable with respect to  $(\mathcal{F}_t)_{t \geq 0}$ . Since  $D - A$  is a martingale, we have  $\mathbb{E}(D(T)) = \mathbb{E}(A(T))$  for any bounded stopping time  $T$ . We conclude that for every bounded stopping time  $T$ , it holds that  $\mathbb{E}(R(T)) = \mathbb{E}(A(T))$ , i.e.,  $A$  majorises  $R$ . Then, if  $A(\tau_n) = o_p(1)$ , the Lenglart-Rebolledo inequality concludes that  $R(\tau_n) = o_p(1)$ , and if  $A(\tau_n) = O_p(1)$ , then the Lenglart-Rebolledo inequality yields  $R(\tau_n) = O_p(1)$ .

## E.2 Proof of Proposition 23

**Item i:** Observe that

$$\tilde{\mathbb{E}}(J(s_i, s_i)) = \tilde{\mathbb{E}}\left(\int_0^{X_i} \int_0^{X_i} A(x, y)dM^i(x)dM^i(y)\right) = \tilde{\mathbb{E}}(D(X_i)) + 2\tilde{\mathbb{E}}(Z(X_i)),$$

where  $D(t)$  and  $Z(t)$  are defined by

$$D(t) = \int_0^t A(x, x)d[M^i](x) \quad \text{and} \quad Z(t) = \int_0^t \int_{(0, y)} A(x, y)dM^i(x)dM^i(y),$$

and notice that  $d[M^i](x) = dN^i(x)$  since the survival times are continuous.

We begin by noticing that

$$\tilde{\mathbb{E}}(D(X_i)) = \tilde{\mathbb{E}}\left(\int_0^{X_i} A(x, x)dN^i(x)\right) = \tilde{\mathbb{E}}(A(X_i, X_i)\Delta_i) = \int_0^\tau A(x, x)(1 - G_{c_i})dF_{c_i}(x).$$

Also, by Theorem 15 and conditioned on the value of the covariate,  $Z(t)$  is a zero-mean  $(\mathcal{F}_t)$ -martingale. Then the optional stopping time theorem yields  $\tilde{\mathbb{E}}(Z(X_i)) = \tilde{\mathbb{E}}(Z(0)) = 0$ , and thus

$$\tilde{\mathbb{E}}(J(s_i, s_i)) = \int_0^\tau A(x, x)(1 - G_{c_i})dF_{c_i}(x).$$

**Item ii:** Observe that

$$|J(s_i, s_i)| \leq \Delta_i |A(X_i, X_i)| + 2\Delta_i \left| \int_0^{X_i} A(X_i, x)d\Lambda_{c_i}(x) \right| + \left| \int_0^{X_i} \int_0^{X_i} A(x, y)d\Lambda_{c_i}(x)d\Lambda_{c_i}(y) \right|.$$

We continue by proving that all the terms on the right-hand-side of the previous equation have positive expectation.

For the first term we have

$$\tilde{\mathbb{E}}(\Delta_i |A(X_i, X_i)|) = \int_0^\tau |A(x, x)|(1 - G_{c_i}(x))dF_{c_i}(x) < \infty, \quad (60)$$

where the last inequality follows from the assumptions made in the Proposition. The second term satisfies

$$\tilde{\mathbb{E}}\left(\Delta_i \left| \int_0^{X_i} A(X_i, x)d\Lambda_{c_i}(x) \right|\right) \leq \int_0^\tau \int_0^y |A(y, x)|d\Lambda_{c_i}(x)(1 - G_{c_i}(y))dF_{c_i}(y) < \infty, \quad (61)$$

where the last inequality follows from Lemma 1 of Efron and Johnstone [1990] (see also Remark 2 of Akritas [2000]). Finally, the last term satisfies

$$\begin{aligned} \tilde{\mathbb{E}}\left(\left| \int_0^{X_i} \int_0^{X_i} A(x, y)d\Lambda_{c_i}(x)d\Lambda_{c_i}(y) \right|\right) &\leq \tilde{\mathbb{E}}\left(\int_0^\tau \int_0^\tau \mathbf{1}_{\{X_i \geq \max(x, y)\}} |A(x, y)|d\Lambda_{c_i}(x)d\Lambda_{c_i}(y)\right) \\ &= \int_0^\tau \int_0^\tau |A(x, y)|(1 - H_{c_i}(x \vee y))d\Lambda_{c_i}(x)d\Lambda_{c_i}(y) \\ &= 2 \int_0^\tau \int_0^y A(x, y)(1 - G_{c_i}(y))dF_{c_i}(x)d\Lambda_{c_i}(y) \\ &< \infty, \end{aligned}$$



where the second equality holds by the symmetry of the functions  $A(x, y)$  and  $1 - H_{c_i}(x \vee y)$ , and the last inequality holds by Equation (61).

**Item iii:** Let  $i \neq j$  and define the process

$$Q(t) = \int_0^t \int_0^{X_j} A(x, y) dM^j(y) dM^i(x). \quad (62)$$

Notice that conditioned on  $s_j = (X_j, \Delta_j, c_j)$  and  $c_i$ , the process  $(Q(t))_{t \geq 0}$  is a zero-mean martingale. Therefore, by the optional stopping theorem we have

$$\tilde{\mathbb{E}}(J(s_i, s_j) | s_j) = \tilde{\mathbb{E}}(Q(X_i) | s_j) = \tilde{\mathbb{E}}(Q(0) | s_j) = 0.$$

**Item iv:** Observe that  $\tilde{\mathbb{E}}(J(s_i, s_j)^2) = \tilde{\mathbb{E}}(Q(X_i)^2)$  for  $i \neq j$ , where  $(Q(t))_{t \geq 0}$  is the process defined in Equation (62). Observe that, conditioned on  $s_j = (X_j, \Delta_j, c_j)$  and  $c_i$ ,  $(Q(t)^2)_{t \geq 0}$  is an  $(\mathcal{F}_t)$ -submartingale with compensator given by

$$\langle Q \rangle(X_i) = \int_0^{X_i} \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 Y^i(x) d\Lambda_{c_i}(x).$$

Then

$$\begin{aligned} \tilde{\mathbb{E}}(Q(X_i)^2 | s_j) &= \tilde{\mathbb{E}}(\langle Q \rangle(X_i) | s_j) = \tilde{\mathbb{E}} \left( \int_0^{X_i} \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 Y^i(x) d\Lambda_{c_i}(x) \middle| s_j \right) \\ &= \int_0^\tau \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 \mathbb{E}(Y^i(x)) d\Lambda_{c_i}(x) \\ &= \int_0^\tau \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 (1 - G_{c_i}(x)) dF_{c_i}(x) \end{aligned}$$

and thus

$$\begin{aligned} \tilde{\mathbb{E}}(Q(X_i)^2) &= \tilde{\mathbb{E}}(\tilde{\mathbb{E}}(Q(X_i)^2 | s_j)) = \tilde{\mathbb{E}} \left( \int_0^\tau \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 (1 - G_{c_i}(x)) dF_{c_i}(x) \right) \\ &= \int_0^\tau \tilde{\mathbb{E}} \left( \left( \int_0^{X_j} A(x, y) dM^j(y) \right)^2 \right) (1 - G_{c_i}(x)) dF_{c_i}(x) \\ &= \int_0^\tau \tilde{\mathbb{E}} \left( \int_0^\tau A(x, y)^2 Y_j(y) d\Lambda_{c_j}(y) \right) (1 - H_{c_i}(x)) \Lambda_{c_i}(x) \\ &= \int_0^\tau \int_0^\tau A(x, y)^2 (1 - G_{c_j}(y)) (1 - G_{c_i}(x)) dF_{c_j}(y) dF_{c_i}(x), \end{aligned}$$

where the second equality is due to the independence of  $s_i$  and  $s_j$  for  $i \neq j$ , and the third equality follows by noticing that, conditioned on  $c_j$ ,  $\left( \int_0^t A(x, y) dM^j(y) \right)^2$  is an  $(\mathcal{F}_t)$ -submartingale for any fixed  $x \in \mathbb{R}_+$  and its compensator is given by  $\int_0^t A(x, y)^2 Y_j(y) d\Lambda_{c_j}(y)$ .