A kernel log-rank test of independence for right-censored data

Tamara Fernández*
University College London
t.a.fernandez.aguilar@gmail.com

Arthur Gretton
University College London
a.gretton@ucl.ac.uk

David Rindt
University of Oxford
david.rindt@magd.ox.ac.uk

Dino Sejdinovic
University of Oxford
dino.sejdinovic@stats.ox.ac.uk

November 18, 2020

Abstract

We introduce a general non-parametric independence test between right-censored survival times and covariates, which may be multivariate. Our test statistic has a dual interpretation, first in terms of the supremum of a potentially infinite collection of weight-indexed log-rank tests, with weight functions belonging to a reproducing kernel Hilbert space (RKHS) of functions; and second, as the norm of the difference of embeddings of certain finite measures into the RKHS, similar to the Hilbert-Schmidt Independence Criterion (HSIC) test-statistic. We study the asymptotic properties of the test, finding sufficient conditions to ensure our test correctly rejects the null hypothesis under any alternative. The test statistic can be computed straightforwardly, and the rejection threshold is obtained via an asymptotically consistent Wild Bootstrap procedure. Extensive simulations demonstrate that our testing procedure generally performs better than competing approaches in detecting complex non-linear dependence.

1 Introduction

Right-censored data appear in survival analysis and reliability theory, where the time-to-event variable one is interested in modelling may not be observed fully, but only in terms of a lower bound. This is a common occurrence in clinical trials as, usually, the follow-up is restricted to the duration of the study, or patients may decide to withdraw from the study.

An important task when dealing with such data is to test independence between the survival times and the covariates. For instance, in a clinical trial setting, we may wish to test if the survival times differ across treatments, e.g., chemotherapy vs radiation, ages of the patients, gender, or any other measured variables. The main challenge of testing independence in this setting is that we need to deal with censored observations, and that the censoring mechanism may be dependent on the covariates while the time of interest may not. For example, patients' withdrawal times from a study can be associated to their gender even if gender is independent of the survival time.

The problem of testing independence has been widely studied by the statistical community. In the context of right-censored data, this problem has often been addressed through the mechanism of two-sample tests, in which the covariate takes one out of two possible values. For the two-sample problem, the main tool is the log-rank test [23, 26] and its generalizations, namely weighted log-rank tests [20, 2, 9]. The more general case, in which the covariates belong to $\mathbb{R}^d$, is much more challenging, and most of the current approaches are ad-hoc for specific semi-parametric models, e.g. [5, 14, 34]. In particular, the most popular of these approaches is the Cox-proportional hazards model [5], which assumes a linear effect of the covariates on the log-hazard function. Non-parametric approaches are more scarce, however [22, 24, 27]. In [22], a nonparametric test for independence is obtained by measuring monotonic relationships between a censored survival time and an ordinal covariate; and in [24], the authors propose an omnibus test that can detect any type of association between a censored survival and a 1-dimensional covariate. The recent non-parametric test in [27] was introduced to deal with general covariates on $\mathbb{R}^d$. This approach deals with censored data by transforming it into uncensored samples,

*corresponding author

and then applies a well-known kernel independence test based on the Hilbert-Schmidt Independence Criterion (HSIC) [15, 16, 4].

In this paper we propose a non-parametric test which can potentially detect any type of dependence between right-censored survival times and general covariates. Our testing procedure is based on a dependence measure between survival times and covariates which is constructed using weighted log-rank tests and the theory of reproducing kernel Hilbert spaces (RKHSs). We provide asymptotic results for our test-statistic and propose an approximation of the rejection region of our test by using a Wild Bootstrap procedure. Under mild regularity conditions, we prove that both the oracle and the testing procedure based on the Wild Bootstrap approximations are asymptotically consistent, meaning that our test can detect any type of dependence.

The closest prior works to our approach are [9] and [27], which are both based on kernel methods. In the former, the authors specifically address the two-sample problem for right-censored data. While the present test may be seen as related, the two-sample analysis is quite different, and heavily relies on the binary nature of the covariates: the main results apply ad-hoc theory developed for log-rank tests, which is not available in our setting. As noted above, [27] bypass the problem of right-censored data by transforming it into uncensored samples, however this comes at the cost of losing considerable information on the data. By contrast, our approach deals directly with the censored observations without loss of information, resulting in a major performance advantage in practice, as we demonstrate in our experiments.

The paper is structured as follows. In Section 2 we introduce relevant notation. In Section 3 we define the kernel log-rank test and show that it can be interpreted as i) the supremum of a collection of score tests associated to a particular family of cumulative hazard functions, and, ii) as an RKHS distance, revealing a similarity with the Hilbert Schmidt independence criterion (HSIC) [18]. In Section 4 we study the asymptotic behavior of our statistic under both the null and alternative hypothesis. We also establish connections with known approaches such as the two-sample test proposed in [9], and the Cox-Score test. These relationships follow by choosing a particular kernel function. Section 5 shows how to effectively approximate the null distribution by using a Wild Bootstrap sampling procedure. Section 6 contains extensive experiments where we evaluate the proposed family of tests for different choices of kernels, and compare the performance with the Cox-score test and the optHSIC test of [27].

2 Notation

Survival Analysis notation Let $((Z_i, C_i, X_i))_{i\in[n]}$ be a collection of random variables taking values on $\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d$, where Z_i denotes a survival time of interest, C_i is a censoring time, and X_i is a vector of covariates taking values on $\mathbb{R}^d$, and $d \geq 1$. In practice, we do not observe (Z_i, C_i, X_i) directly, but instead observe triples (T_i, Δ_i, X_i), where $T_i = \min\{Z_i, C_i\}$ and $\Delta_i = \mathbb{1}_{\{T_i=Z_i\}}$. This type of data is known as right-censored data. Additionally, we assume $Z \perp C|X$, which is known as independent right censoring.

We denote by F_T, F_Z, F_C and F_X, the marginal distribution functions associated to T, Z, C and X, respectively. We use standard notation to denote joint and conditional distributions, e.g. $F_{ZC|X=x}$ denotes the joint distribution of Z and C conditional on $X = x$. We denote by $S_T(t) = 1 - F_T(t)$, $S_Z(t) = 1 - F_Z(t)$ and $S_C(t) = 1 - F_C(t)$ the marginal survival functions associated to T, Z and C, respectively, and by $S_{T|X=x}(t) = 1 - F_{T|X=x}(t)$, $S_{Z|X=x}(t) = 1 - F_{Z|X=x}(t)$ and $S_{C|X=x}(t) = 1 - F_{C|X=x}(t)$, the respective survival functions conditioned on $X = x$. In this work, we assume that $Z|X = x$ and $C|X = x$ are continuous random variables for almost all $x \in \mathbb{R}^d$, with densities denoted by $dF_{Z|X=x}(t)$ and $dF_{C|X=x}(t)$ respectively. We further assume that Z and C are proper random variables, meaning that $\mathbb{P}(Z < \infty|X = x) = 1$ and $\mathbb{P}(C < \infty|X = x) = 1$ for almost all $x \in \mathbb{R}^d$. The marginal cumulative hazard function of Z is defined as $\Lambda_Z(t) = \int_0^t S_Z(s)^{-1}dF_Z(s)$. Similarly, the conditional cumulative hazard of Z given $X = x$ is $\Lambda_{Z|X=x}(t) = \int_0^t S_{Z|X=x}(s)^{-1}dF_{Z|X=x}(s)$. We define $\tau_n = \max\{T_1, \ldots, T_n\}$, $\tau_x = \sup\{t : S_{T|X=x}(t) > 0\}$ and $\tau = \sup\{t : S_T(t) > 0\}$; note that $\tau_n \overset{a.s.}{\to} \tau$.

Counting processes notation We use standard Survival Analysis/counting processes notation. For $i \in [n]$, we define the individual and pooled counting processes by $N_i(t) = \Delta_i \mathbb{1}_{\{T_i\leq t\}}$ and $N(t) = \sum_{i=1}^n N_i(t)$, respectively. Similarly, we define the individual and pooled risk functions by $Y_i(t) = \mathbb{1}_{\{T_i\geq t\}}$ and $Y(t) = \sum_{i=1}^n Y_i(t)$.

We assume that all our random variables take values on a common filtrated probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where the sigma-algebra $\mathcal{F}_t$ is generated by

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

and the $\mathbb{P}$-null sets of $\mathcal{F}$. Under the null hypothesis, for $i \in [n]$, we define the individual and pooled $(\mathcal{F}_t)$-martingales, $M_i(t) = N_i(t) - \int_{(0,t]} Y_i(s)d\Lambda_Z(s)$ and $M(t) = N(t) - \int_{(0,t]} Y(s)d\Lambda_Z(s)$, respectively. Finally, we denote by $d\widehat{\Lambda}(t) = dN(t)/Y(t)$ the Nelson Aalen estimator of $d\Lambda_Z(t)$ under the null hypothesis. For more information about counting processes martingales, we refer the reader to Fleming and Harrington [10, Chapters 1 and 2].

In this work $\int_a^b$ means integration over $(a, b]$ unless $b = \tau$, in which case we integrate over (a, τ). Due to the simple nature of the martingales that appear in this work (which arise from counting processes) properties such as integrability or squared-integrability of these processes are standard, and thus we state them without formal proof. Also, note that for any $t > \tau_n$, it holds that $N(t) = N(\tau_n)$ and $M(t) = M(\tau_n)$. Hence $\int_{\mathbb{R}_+} g(t)dN(t) = \int_0^\tau g(t)dN(t) = \int_0^{\tau_n} g(t)dN(t)$; the same holds for the martingale M.

For simplicity of exposition and notation we assume $X \in \mathbb{R}^d$, however our results also apply straightforwardly to general covariate spaces, as our statistic is based on kernel functions that may be defined on more general domains: see next section.

3 Construction of the test

We are interested in testing if the failure times Z are independent of the covariates X. Specifically, we would like to test the null hypothesis,

$$H_0 : F_{ZX} = F_Z F_X, \quad \text{against} \quad H_1 : F_{ZX} \neq F_Z F_X.$$

One of the most popular approaches to solve this problem is the log-rank test for proportional hazard functions. This test can be obtained as a score test from a partial likelihood function for the Cox's proportional hazards model given by $\Lambda_{Z|X=x}(t) = e^{\beta^\intercal x}\Lambda_Z(t)$. This approach fails in many scenarios, however as it only considers a linear effect of the covariates on the log hazard, which is given by the term $\beta^\intercal x$.

Our method generalizes the previous method by defining a general collection of log-rank tests in which the association between time and covariates is modeled through general functions $\omega(t, x)$, instead of the simple expression $\beta^\intercal x$.

General score test We obtain a general log-rank test, for a fixed function $\omega : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, by computing the score test associated to the model defined in terms of the conditional cumulative hazard function,

$$\Lambda_{Z|X=x}(t; \theta, \omega) = \int_0^t e^{\theta\omega(s,x)} d\Lambda_Z(s) \quad \theta \in \Theta, \tag{1}$$

where $\omega : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ is some non-zero fixed function, Θ is an open subset of $\mathbb{R}$ containing $\theta = 0$, and $\Lambda_Z(s)$ is the marginal (baseline) cumulative hazard function associated to the failure time Z. Under the assumption that our data is generated by this model for some fixed function $\omega(t, x)$, testing the null hypothesis $H_0 : Z \perp X$ is equivalent to testing $H_0 : \theta = 0$, which can be done using a score test.

A score test is a hypothesis test used to check whether a restriction imposed on a model estimated by maximum likelihood is violated by the data. The score test assesses the gradient of the log-likelihood function, known as score function, evaluated at some parameter θ under the null hypothesis. Intuitively, if the maximizer of the log-likelihood function is close to 0, the score, evaluated at $\theta = 0$, should not differ from zero by more than sampling error.

The likelihood function associated to the right-censored data $(T_i, \Delta_i, X_i)_{i=1}^n$ can be computed as follows. Given X_i, the contribution to the likelihood of an uncensored observation (T_i, Δ_i), that is, for $\Delta_i = 1$, is $dF_{Z|X_i}(T_i) = d\Lambda_{Z|X_i}(T_i)S_{Z|X_i}(T_i)$. When (T_i, Δ_i) is censored, $\Delta_i = 0$, the contribution corresponds to $S_{Z|X_i}(T_i)$. The latter follows from the fact that when T_i is censored, we only know that Z_i is greater than T_i.

Thus, given the covariates $(X_i)_{i=1}^n$, the likelihood function for the data $(T_i, \Delta_i)_{i=1}^n$ under the model in Equation (1) corresponds to

$$L_n(\theta; \omega) = \prod_{i=1}^n d\Lambda_{Z|X_i}(T_i)^{\Delta_i} S_{Z|X_i}(T_i) = \prod_{i=1}^n e^{\theta\Delta_i\omega(T_i,X_i)} d\Lambda_Z(T_i)^{\Delta_i} \exp\left\{-\int_0^{T_i} e^{\theta\omega(s,X_i)} d\Lambda_Z(s)\right\},$$

where the second equality follows by noticing that $S_{Z|X}(t) = \exp\left\{-\int_0^t d\Lambda_{Z|X}(s)\right\}$.

The score function is then defined as

$$U_n(\theta; \omega) = \frac{d}{d\theta} \log L_n(\theta; \omega) = \sum_{i=1}^n \left(\Delta_i\omega(T_i, X_i) - \int_0^{T_i} \omega(t, X_i)e^{\theta\omega(t,X_i)} d\Lambda_Z(t)\right),$$

and $U_n(0,\omega)$ is the score statistic associated to the null hypothesis, $H_0 : \theta = 0$. A normalized version of $U_n(0,\omega)$ can be obtained using the variance/covariance matrix of $U_n(\theta;\omega)$, written as $\Sigma(\theta;\omega) = \mathbb{E}(-\frac{\partial^2}{\partial\theta^2}\log L_n(\theta;\omega))$, and then writing $S_n(0;\omega) = U_n(0;\omega)^\intercal\Sigma(0;\omega)^{-1}U_n(0;\omega)$. By the Neyman-Pearson Lemma [25], it follows that the test based on $S_n(0;\omega)$ is the most powerful test for small deviations from the null under the model defined in Equation (1).

In general the marginal hazard function $d\Lambda_Z(s)$ is unknown, and thus $U_n(0;\omega)$ can not be evaluated in practice. However, under the null, $d\Lambda_Z(s)$ can be estimated from the data using the Nelson-Aalen estimator [1] $d\widehat{\Lambda}_Z(t) = dN(t)/Y(t)$, giving the expression

$$\widehat{U}_n(0;\omega) = \sum_{i=1}^n \int_{\mathbb{R}_+} (\omega(t,X_i) - \bar{\omega}_n(t))dN_i(t), \tag{2}$$

where $\bar{\omega}_n(t) = \sum_{j=1}^n \omega(t,X_j)Y_j(t)/Y(t)$.

Log-rank formulation The expression for the un-normalized score statistic given in Equation (2) can be written as a discrepancy between two empirical measures with respect to the weight function $\omega(t,x)$. In Survival Analysis terminology, this is known as weighted log-rank test, and, in our scenario, it takes the form

$$\mathrm{LR}_n(\omega) = \frac{1}{n}\widehat{U}_n(0;\omega) = \int_{\mathbb{R}_+}\int_{x\in\mathbb{R}^d} \omega(t,x)\left(d\nu_1^n(t,x) - d\nu_0^n(t,x)\right), \tag{3}$$

where ν_1^n and ν_0^n are empirical measures defined as

$$d\nu_1^n(t,x) = \frac{1}{n}\sum_{i=1}^n dN_i(t)\delta_{X_i}(x) = \frac{1}{n}\sum_{i=1}^n \Delta_i\delta_{T_i,X_i}(t,x) \tag{4}$$

and

$$d\nu_0^n(t,x) = \frac{dN(t)}{n}\sum_{i=1}^n \frac{Y_i(t)}{Y(t)}\delta_{X_i}(x) = \frac{1}{n}\sum_{j=1}^n \Delta_j\delta_{T_j}(t)\sum_{i=1}^n \frac{Y_i(t)}{Y(t)}\delta_{X_i}(x). \tag{5}$$

The next theorem gives a consistency limit result for $\mathrm{LR}(\omega)$.

Theorem 3.1. *Let $\omega : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function. Then*

$$LR_n(\omega) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)(d\nu_1(s,x) - d\nu_0(s,x)),$$

where $d\nu_1(t,x) = S_{C|X=x}(t)dF_{ZX}(t,x)$, $d\nu_0(t,x) = S_{T|X=x}(t)d\alpha(t)dF_X(x)$, and the measure α is defined as $\alpha(I) = \int_I\int_{\mathbb{R}^d} S_{C|X=x}(t)/S_T(t)dF_{ZX}(t,x)$ for any measurable $I \subseteq (0,\tau)$.

Simple algebra shows that, under the null hypothesis (i.e., $H_0 : Z \perp X$), $\nu_1 = \nu_0$, and consequently $\mathrm{LR}_n(\omega) \xrightarrow{\mathbb{P}} 0$ for any weight function $\omega(t,x)$. Under some regularity conditions which we state in Assumption 3.2, we prove $\nu_1 = \nu_0$ implies $Z \perp X$.

Assumption 3.2. *Assume that $S_{C|X=x}(t) = 0$ implies $S_{Z|X=x}(t) = 0$ for almost all $x \in \mathbb{R}^d$.*

Proposition 3.3. *Under Assumption 3.2, it holds that $\nu_1 = \nu_0$ if and only if $Z \perp X$.*

Note that $\mathrm{LR}_n(\omega) \xrightarrow{\mathbb{P}} 0$ does not necessarily imply $\nu_0 = \nu_1$, since, if we choose ω equal to the zero function, then $\mathrm{LR}_n(\omega) = 0$ trivially. Thus, when using log-rank tests, it is very important to use a relevant weight function for the problem at hand.

Instead of choosing a single weight function $\omega(t,x)$, our approach will optimize over a large collection of candidate functions.

RKHS approach While normalized log-rank tests exhibit good statistical properties for small deviations from alternatives belonging to the model in Equation (1), this good behavior is only guaranteed for a single weight function ω at a time. In practice, it is very unlikely to know beforehand the dependence structure of Z and X, and thus choosing the correct weight $\omega(t,x)$ (if it exists) seems hard.

In order to avoid choosing a particular weight $\omega(t,x)$, we consider a family of weighted log-rank statistics, and compute

$$\Psi_n^2 = \left(\sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \leq 1} \mathrm{LR}_n(\omega) \right)^2, \tag{6}$$

where the function $\omega(t,x)$ is allowed to take values in a potentially infinite-dimensional space of functions $\mathcal{H}$. We refer to Ψ_n^2 as the *kernel log-rank* statistic.

In particular, we choose $\mathcal{H}$ as a reproducing kernel Hilbert space (RKHS) of functions. One of the main advantages of choosing this particular space of functions, is that it gives a simple close-form solution for the optimization problem of Equation (6). For general spaces of functions, finding the function $\widehat{\omega}$ that maximizes the likelihood function, or solving the optimization problem of Equation (6) might be much harder problem, as it is likely that $\widehat{\omega}$ does not have a close-form solution. While our approach seems ad-hoc, we will prove that, under some mild regularity assumptions, this space of functions is rich enough to be able to detect any type of dependencies. As opposed to other works that consider maximum among normalised log-rank statistics (i.e., divided by the standard deviation) [21, 31, 11], our statistic uses the un-normalised version $\mathrm{LR}_n(\omega)$. This is fundamental in our results as the linearity in ω of $\mathrm{LR}_n(\omega)$, combined with the properties of the RKHS, lead to a simple closed formula to evaluate Ψ_n^2. This being said, note that we are indirectly normalizing by choosing ω in the unit ball of $\mathcal{H}$.

Reproducing Kernel Hilbert Spaces An RKHS $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ is a Hilbert space of functions in which the evaluation operator is continuous. By the Riesz representation theorem, for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ there exists an unique element $\mathfrak{K}_{(t,x)} \in \mathcal{H}$ such that, for all $\omega \in \mathcal{H}$ it holds $\omega(t,x) = \langle \omega, \mathfrak{K}_{(t,x)} \rangle_{\mathcal{H}}$; this property is known as the *reproducing property*. We define the so-called reproducing kernel $\mathfrak{K} : (\mathbb{R}_+ \times \mathbb{R}^d)^2 \to \mathbb{R}$ as $\mathfrak{K}((t,x),(t',x')) = \langle \mathfrak{K}_{(t',x')}, \mathfrak{K}_{(t,x)} \rangle_{\mathcal{H}}$ for any $(t,x),(t',x') \in \mathbb{R}_+ \times \mathbb{R}^d$. Note that the reproducing kernel is unique and positive definite. By the Moore-Aronszajn theorem, for any symmetric positive-definite kernel $\mathfrak{K}$, there exists a unique RKHS for which $\mathfrak{K}$ is its reproducing kernel. Finally, for any given measure (not necessarily a probability measure), we define its embedding into $\mathcal{H}$ as

$$\phi_\nu(\cdot) = \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \mathfrak{K}((t,x),\cdot) d\nu(t,x) \in \mathcal{H}.$$

The existence of ϕ_ν is guaranteed by $\int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \sqrt{\mathfrak{K}((t,x),(t,x))} d\nu(t,x) < \infty$, see [3, Chapter 3] and [29].

RKHS distance We define the embeddings of the empirical measures ν_1^n and ν_0^n, defined in equations (4) and (5), into $\mathcal{H}$ with reproducing kernel $\mathfrak{K} : (\mathbb{R}_+ \times \mathbb{R}^d)^2 \to \mathbb{R}$ as

$$\phi_1^n(\cdot) = \int_0^\tau \int_{\mathbb{R}^d} \mathfrak{K}((t,x),\cdot) d\nu_1^n(t,x) \quad \text{and} \quad \phi_0^n(\cdot) = \int_0^\tau \int_{\mathbb{R}^d} \mathfrak{K}((t,x),\cdot) d\nu_0^n(t,x), \tag{7}$$

respectively. Notice both ϕ_1^n and ϕ_0^n are well-defined elements of $\mathcal{H}$, as they are just finite sums of elements of $\mathcal{H}$.

The next Theorem gives a closed-form expression for the kernel log-rank statistic in terms of the distance (induced by the norm) of the embeddings ϕ_0^n and ϕ_1^n.

Theorem 3.4.

$$\Psi_n^2 = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \Delta_i \Delta_j \bar{\mathfrak{K}}_n((Z_i,X_i),(Z_j,X_j)), \tag{8}$$

where

$$\begin{aligned} \bar{\mathfrak{K}}_n((t,x),(t',x')) = \mathfrak{K}((t,x),(t',x')) &- \sum_{k=1}^n \mathfrak{K}((t,x),(t',X_k)) \frac{Y_k(t')}{Y(t')} \\ &- \sum_{j=1}^n \mathfrak{K}((t,X_j),(t',x')) \frac{Y_j(t)}{Y(t)} + \sum_{j,k=1}^n \mathfrak{K}((t,X_j),(t',X_k)) \frac{Y_j(t)}{Y(t)} \frac{Y_k(t')}{Y(t')}. \end{aligned}$$

Moreover, if $\mathfrak{K}((t,x),(t',x')) = L(t,t')K(x,x')$, then

$$\Psi_n^2 = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 = \frac{1}{n^2}\operatorname{trace}(\boldsymbol{L}^{\boldsymbol{\Delta}}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{K}(\boldsymbol{I}-\boldsymbol{A})^\intercal) \tag{9}$$

where $\boldsymbol{K}, \boldsymbol{L}^{\boldsymbol{\Delta}}$ and $\boldsymbol{A}$ are $(n\times n)$-dimensional matrices whose entries (i,j) are defined as $(\boldsymbol{K})_{i,j} = K(X_i,X_j)$, $(\boldsymbol{L}^{\boldsymbol{\Delta}})_{i,j} = \Delta_i\Delta_j L(T_i,T_j)$ and $(\boldsymbol{A})_{i,j} = A_{ij} = \frac{Y_j(T_i)}{Y(T_i)}$, and $\boldsymbol{I}$ denotes the identity matrix.

4 Asymptotic Analysis

Asymptotic null distribution We study the asymptotic null distribution of $n\Psi_n^2$ which is fundamental to construct a testing procedure. The key step is to show that we can rewrite Ψ_n^2 as a V-statistic plus an asymptotically negligible term. The asymptotic null distribution of $n\Psi_n^2$ then follows from the standard theory of V-statistics. We refer to [28, Section 5.5.2] for a discussion of V-statistics.

Proposition 4.1. *Under the null hypothesis $H_0 : Z \perp X$, the kernel log-rank statistic can be written as*

$$\Psi_n^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n J_n((T_i,\Delta_i,X_i),(T_j,\Delta_j,X_j)), \tag{10}$$

where $J_n : (\mathbb{R}_+ \times \{0,1\}\times\mathbb{R}^d)^2 \to \mathbb{R}$ is a symmetric random function defined as

$$J_n((s,c,x),(s',c',x')) = \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \bar{\mathfrak{K}}_n((t,x),(t',x'))dm_{s,c}(t)dm_{s',c'}(t),$$

and $dm_{s,c}(t) = c\delta_s(t) - \mathbb{1}_{\{s\geq t\}}d\Lambda_Z(t)$.

The expression in Equation (10) suggests a V-statistic representation for the kernel log-rank statistic. In the next result, we prove that $n\Psi_n^2$ can, indeed, be approximated by a V-statistic, by showing that J_n can be replaced by its population version J, which follows from replacing the random kernel $\bar{\mathfrak{K}}_n$ by its corresponding population version $\bar{\mathfrak{K}}$, given by

$$\begin{aligned}\bar{\mathfrak{K}}((t,x),(t',x')) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Bigg(&\mathfrak{K}((t,x),(t',x')) - \mathfrak{K}((t,x),(t',y'))\frac{S_{C|X=y'}(t')}{S_C(t')}\\ &- \mathfrak{K}((t,y),(t',x'))\frac{S_{C|X=y}(t)}{S_C(t)} + \mathfrak{K}((t,y),(t',y'))\frac{S_{C|X=y}(t)S_{C|X=y'}(t')}{S_C(t)S_C(t')}\Bigg)dF_X(y)dF_X(y'),\end{aligned} \tag{11}$$

which is valid under the null as $S_T(t) = S_Z(t)S_C(t)$.

Assumption 4.2. $\mathfrak{K}((t,x),(t',x')) = L(t,t')K(x,x')$*, and both K and L are bounded.*

Lemma 4.3. *Under Assumption 4.2 and the null hypothesis, it holds*

$$\Psi_n^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n J((T_i,\Delta_i,X_i),(T_j,\Delta_j,X_j)) + o_p(n^{-1}),$$

where

$$J((s,c,x),(s',c',x')) = \int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\bar{\mathfrak{K}}((t,x),(t',x'))dm_{s,c}(t)dm_{s',c'}(t). \tag{12}$$

It can be easily checked that $\mathbb{E}(J((t,c,x),(T_1,\Delta_1,X_1))) = 0$ for any $(t,c,x)\in\mathbb{R}_+\times\{0,1\}\times\mathbb{R}^d$ under the null (since $dm_{T_i,\Delta_i}(t) = dM_i(t)$). The statistic Ψ_n^2 is then approximately a degenerate V-statistic, and thus we deduce its limit distribution from the classical theory of degenerate V-statistics [28, Section 5.5.2].

Theorem 4.4. *Under Assumption 4.2 and the null hypothesis, it holds that*

$$n\Psi_n^2 \overset{\mathcal{D}}{\to} \int_{x\in\mathbb{R}^d}\int_0^\tau \bar{\mathfrak{K}}((t,x),(t,x))S_{C|X=x}(t)dF_Z(t)dF_X(x) + \mathcal{Y}$$

where $\mathcal{Y} = \sum_{i=1}^\infty \lambda_i(\xi_i^2-1)$, $\xi_1,\xi_2,\dots$ are i.i.d. standard normal random variables, and $\lambda_1,\lambda_2,\dots$ are non-negative constants which depend on the distribution of the random variables (Z,C,X) and the kernel $\mathfrak{K}$.

The next result states that if we directly replace $\bar{\mathfrak{K}}_n$ by its limit $\bar{\mathfrak{K}}$ in Equation (8), the resulting test-statistic has the same asymptotic null distribution as $n\Psi_n^2$. This result will be important in Section 5 to show that the asymptotic distribution of the Wild Bootstrap test-statistic is the same as the asymptotic null distribution of the kernel log-rank statistic.

Theorem 4.5. *$n\Psi_n^2$ and $\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \Delta_i\Delta_j\bar{\mathfrak{K}}((Z_i,X_i),(Z_j,X_j))$ have the same asymptotic distribution under the null hypothesis.*

Power under alternatives We next analyze the asymptotic behavior of Ψ_n^2 under the alternative hypothesis, i.e., $H_1 : F_{ZX} \neq F_Z F_X$. To this end, we first establish a consistency result for Ψ_n^2.

Lemma 4.6. *Under Assumption 4.2, it holds $\Psi_n^2 \overset{\mathbb{P}}{\to} \|\phi_0 - \phi_1\|_{\mathcal{H}}^2$, where*

$$\phi_1(\cdot) = \int_0^\tau \int_{\mathbb{R}^d} K(\cdot,x)L(\cdot,t)d\nu_1(t,x) \quad \textit{and} \quad \phi_0(\cdot) = \int_0^\tau \int_{\mathbb{R}^d} K(\cdot,x)L(\cdot,t)d\nu_0(t,x),$$

and ν_1 and ν_0 are the population measures defined in Theorem 3.1.

The next step is to ensure that $\|\phi_0 - \phi_1\|_{\mathcal{H}}^2$ is zero if and only if the null hypothesis holds. This result will follow from assuming conditions on the kernel $\mathfrak{K}$ that ensure the embeddings of the measures ν_0 and ν_1 onto $\mathcal{H}$ are injective, and from Proposition 3.3, which proves $\nu_0 = \nu_1$ if and only if the null hypothesis holds.

Theorem 4.7. *Consider Assumptions 3.2 and 4.2, and assume that both L and K, are characteristic [30], translation invariant, and c_0-kernels. Then, $n\Psi_n^2 \to \infty$ under the alternative hypothesis, and thus*

$$\mathbb{P}\left(n\Psi_n^2 > Q_n^{1-\alpha}\right) \to 1, \qquad \textit{as } n \to \infty,$$

where $\alpha \in (0,1)$, and $Q_n^{1-\alpha}$ is the $1-\alpha$ quantile of the distribution of $n\Psi_n^2$ under the null hypothesis.

In the previous result we say K is a c_0-kernel if $K(x,\cdot) \in \mathcal{C}_0(\mathbb{R}^d)$, where $\mathcal{C}_0(\mathbb{R}^d)$ denotes the class of continuous functions in $\mathbb{R}^d$ that vanish at infinity. An example of a kernel that satisfies the conditions stated in the previous Theorem, and that satisfies Assumption 4.2, is the exponential quadratic kernel, given by $K(x,y) = \exp\{-(x-y)^\intercal\Sigma^{-1}(x-y)\}$.

Under the assumptions of Theorem 4.7, the oracle testing procedure, which is based on the $1-\alpha$ quantile of $n\Psi_n^2$, has asymptotic power tending to one for any alternative, and thus *it is able to detect any type of dependency* between survival times and covariates with enough observations. Even if the kernel does not satisfy the properties stated in Theorem 4.7, however, we can guarantee the power of the oracle test for alternatives following the model of Equation (1), that is, for alternatives of the form $\Lambda_{Z|X=x}(t;\theta) = \int_0^t e^{\theta\omega^\star(s,x)}d\Lambda_Z(s)$ for some $\omega^\star \in \mathcal{H}$ on the unit ball and $\theta \neq 0$, as the log-rank statistic $\mathrm{LR}_n(\omega^\star)^2 \to c > 0$, with c a positive constant. Then,

$$\Psi_n^2 = \left(\sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}^2\leq 1} \mathrm{LR}_n(\omega)\right)^2 \geq \mathrm{LR}_n(\omega^\star)^2 \to c,$$

when re-scaling by n, it holds that $n\Psi_n^2 \to \infty$.

Recovering existing tests We show our approach can also recover certain known tests for specific choices of the kernel function.

Example 4.8 (Two-sample weighted log-rank test)**.** Consider $X \in \{0,1\}$, i.e., the two-sample problem. We can recover the standard weighted log-rank test with arbitrary weight function $\tilde{\omega} : \mathbb{R}_+ \to \mathbb{R}$, by choosing $\omega(t,1) = -\omega(t,0)$ and $\omega(t,0) = \tilde{\omega}(t)/2$. Then, by replacing ω into Equation (3), we obtain

$$\mathrm{LR}_n(\omega) = \frac{1}{n}\int_0^\tau \tilde{\omega}(t)L(t)(d\widehat{\Lambda}^0(t) - d\widehat{\Lambda}^1(t)), \tag{13}$$

where $d\widehat{\Lambda}^j$ denotes the Nelson-Aalen estimator for each group $j \in \{0,1\}$. Furthermore, $\tilde{\Psi}_n = \sup_{\omega:\|\omega\|_{\mathcal{H}}=1}\mathrm{LR}_n(\omega)$ recovers the general test proposed in [9] for the two-sample problem.

Example 4.9 (Cox proportional hazards model)**.** Consider the Hilbert space of functions $\omega(t,x) = V^{1/2}\beta^\intercal x$, where $\beta \in \mathbb{R}^d$ and V is a positive-definite matrix of length-scales. By using this space of functions, our kernel log-rank statistic becomes

$$\Psi_n = \sup_{\beta\in\mathbb{R}^d:\|V^{1/2}\beta\|^2\leq 1} \mathrm{LR}_n(\beta) \tag{14}$$

and it can be computed using Equation (9) with a linear kernel on the covariates, $K(x,x') = (V^{1/2}x)^\intercal(V^{1/2}x')$, and a constant kernel on times $L(t,t') = 1$. Then

$$n\Psi_n^2 = \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^n \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \bar{\mathfrak{K}}_n((t,X_i),(t',X_l))dM_i(t)dM_l(t') = U_{\mathrm{Cox}}(0)^\intercal V U_{\mathrm{Cox}}(0),$$

where $U_{\mathrm{Cox}}(0) = \frac{d}{d\beta}l_{\mathrm{Cox}}(\beta)\Big|_{\beta=0}$ is the score function associated to the so-called *Cox partial likelihood* $l_{\mathrm{Cox}}(\beta)$. By choosing V equal to the inverse of the Fisher information matrix, the Cox score test and our Ψ_n are asymptotically equivalent.

5 Wild Bootstrap implementation

Let $\alpha \in (0,1)$, and denote by $Q_n^{1-\alpha}$ and $Q^{1-\alpha}$ the $1-\alpha$ quantile of the null distribution of $n\Psi_n^2$ under the finite sample and asymptotic regime, respectively. In practice, excluding exceptional cases, it is not possible to access these quantiles, and thus we propose to use a Wild Bootstrap approximation of the asymptotic null distribution.

The Wild Bootstrap test-statistic is given by

$$(\Psi_n^W)^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n W_iW_j\Delta_i\Delta_j\bar{\mathfrak{K}}_n((Z_i,X_i),(Z_j,X_j)),$$

where $W = (W_1\ldots,W_n)$ are a collection of i.i.d. Rademacher random variables, which are independent of the data $\mathcal{D} = \{(T_i,\Delta_i,X_i)\}_{i=1}^n$. Note that the only difference between Ψ_n^2 and $(\Psi_n^W)^2$ are the Wild Bootstrap weights W_i that appear in $(\Psi_n^W)^2$.

In this section we prove two main results. The first result establishes that the asymptotic distribution of $n(\Psi_n^W)^2$ coincides with the asymptotic distribution of the kernel log-rank test-statistic, $n\Psi_n^2$, under the null hypothesis. The second result is analogous to Theorem 4.7, but for the $1-\alpha$ quantile obtained by the Wild Bootstrap procedure.

Lemma 5.1. *Under Assumption 4.2, it holds*

$$(\Psi_n^W)^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n W_iW_j\Delta_i\Delta_j\bar{\mathfrak{K}}((Z_i,X_i),(Z_j,X_j)) + o_p(n^{-1}). \tag{15}$$

The previous result establishes that we can directly replace the random kernel $\bar{\mathfrak{K}}_n$ by its population version $\bar{\mathfrak{K}}$ (up to a negligible term) without the need to rely on a martingale representation, as we did for the kernel log-rank test-statistic in Proposition 4.1. This is possible due to the presence of Rademacher random variables and the use of *symmetrization inequalities* (e.g., [32, Lemma 6.4.2]). This result, together with Theorem 4.5 and [6, Theorem 3.1], proves our first main result.

Theorem 5.2. *Under Assumption 4.2, the asymptotic distribution of $n(\Psi_n^W)^2$ and the asymptotic distribution of $n\Psi_n^2$ are equal under the null hypothesis.*

Our second main result is given in the following theorem.

Theorem 5.3. *Consider Assumptions 3.2 and 4.2, and assume that both L and K, are characteristic, translation invariant, and c_0-kernels. Let $\alpha \in (0,1)$, and let $Q_{n,M}^W$ denote the $1-\alpha$ quantile obtained from a sample of size M of the Wild Bootstrap test-statistic, $n(\Psi_n^W)_1^2,\ldots,n(\Psi_n^W)_M^2$. Then, under the alternative hypothesis, it holds*

$$\mathbb{P}\left(n\Psi_n^2 > Q_{n,M}^W\right) \to 1 \qquad \textit{as } n\to\infty,$$

for any fixed M.

From the previous result, we deduce that, under the Assumptions of Theorem 5.3, the test based on the Wild Bootstrap approximation of the null distribution is able to *detect any type of dependency between survival times and covariates* asymptotically, as long as censoring does not hide the regions in which the depende occurs.

Implementation Under Assumption 4.2, the Wild Bootstrap test-statistic can be easily evaluated as follows

$$n(\Psi_n^2)^W = \frac{1}{n^2}\,\mathrm{trace}(\boldsymbol{L}^{\boldsymbol{\Delta},\boldsymbol{W}}(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{K}(\boldsymbol{I}-\boldsymbol{A})^\intercal) \tag{16}$$

where $\boldsymbol{L}^{\boldsymbol{\Delta},\boldsymbol{W}}$ is a $(n \times n)$-matrix defined as $(\boldsymbol{L}^{\boldsymbol{\Delta},\boldsymbol{W}})_{i,j} = (\Delta_i W_i)(\Delta_j W_j)L(T_i,T_j)$, and $\boldsymbol{K}, \boldsymbol{A}$ and $\boldsymbol{I}$ are defined in Theorem 3.4. Algorithm 1 below describes the implementation of our testing procedure.

Computational time By the following Proposition, our algorithm has the same computational complexity as the HSIC based permutation test of [18].

Algorithm 1: Wild Bootstrap.

Input: data $\{T_i, \Delta_i, X_i\}_{i=1}^n$, α and M
for *k in* $1 \to M$ **do**
  Sample $W = (W_1, \dots, W_n) \overset{i.i.d.}{\sim}$ Rademacher
  Compute $(\Psi_n^W)_k^2$ as in equation (16)
Denote by $Q_{n,M}^W$ the $1-\alpha$ quantile of the sample $n(\Psi_n^W)_1^2, \dots, n(\Psi_n^W)_M^2$
Compute $n\Psi_n^2$ as in Equation (9)
Reject if $n\Psi_n^2 > Q_{n,M}^W$

Proposition 5.4. *$n\Psi_n^2$ and $n(\Psi_n^W)^2$ can be computed in $\mathcal{O}(n^2)$ time.*

Using a simple Python implementation that does not use a GPU, running on CPU with 4 cores at 1.6GHz, computation of the kernel log-rank statistic takes about 10 seconds for a sample of size 10000, and about 0.1 second for a sample of size 1000. If faster computation is required, we may adopt the large-scale approximations proposed in [33]. Moreover, Wild Bootstrap statistics can be computed in parallel, and matrix computations can be done on a GPU.

6 Experiments

We study the performance of our proposed kernel log-rank test with different choices of kernels. We choose the kernels to be products of a kernel on the covariates, K, and a kernel on the times, L. We denote this by (K, L). We study the following four cases: 1. $(K = \text{Lin}, L = 1)$, 2. $(K = \text{Gau}, L = 1)$, 3. $(K = \text{Fis}, L = 1)$ and 4. $(K = \text{Gau}, L = \text{Gau})$, where "Lin" denotes the linear kernel, "Gau" denotes a Gaussian kernel, "Fis" is the linear kernel scaled by the Fisher information, see Example 4.9 and "1" denotes the constant kernel, i.e. $L = 1$ implies $L(t,s) = 1$ for all $t, s \in \mathbb{R}_+$. In all the experiments we use the median-distance heuristic to select the bandwidth of the Gaussian kernel: we choose $\sigma^2 = \text{median}\{||x_i - x_j||^2 \ : i \neq j\}/2$. We discuss the sensitivity of the test to different choices of bandwidth in Appendix A.6. We set the level of the test to $\alpha = 0.05$ and use Algorithm 1 in Section 5 to perform the test. We use $M = 1999$ Wild Bootstrap samples for the rejection region. We compare the kernel log-rank test with the traditional Cox-score test [5], denoted by Cph in the legends, and the optHSIC test proposed in [27], denoted by Opt in the legends. As in [27], we use the Brownian covariance kernel in optHSIC.

Type I error We simulate data from distributions for which the null hypothesis holds and compare the Type I error rate of the Wild Bootstrap tests with different kernels with the level $\alpha = 0.05$. Exact descriptions of the distributions (*D.1 - D.6*) and tables of the corresponding rejection rates are given in Appendix A.1 of the supplementary material. The distributions include both univariate and multivariate covariates. We let $C \not\perp\!\!\!\perp X$, because it is of particular interest to check that the Type 1 error rate is correct in this more difficult case (where e.g. a permutation test is not valid). The obtained rejection rates show that the kernel log-rank test achieves the correct Type-I error of 0.05 for each kernel, for each sample size and for each scenario. Importantly the rejection rate is correct even in cases in which the censoring distribution depends strongly on the covariate.

Power for 1-dimensional covariates We now assume that the alternative hypothesis holds, that is, $H_1 : Z \not\perp\!\!\!\perp X$, and that X is a 1-dimensional covariate. We study the power of our test for different distributions of F_{ZCX}. In each case, we use an exponential censoring distribution in which $C \perp\!\!\!\perp X$ and the mean of C is chosen such that 60% of the events are observed. Throughout this subsection we let $X \sim \text{Unif}[-1,1]$.

D.7: Cox proportional hazards: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{x/3\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$.

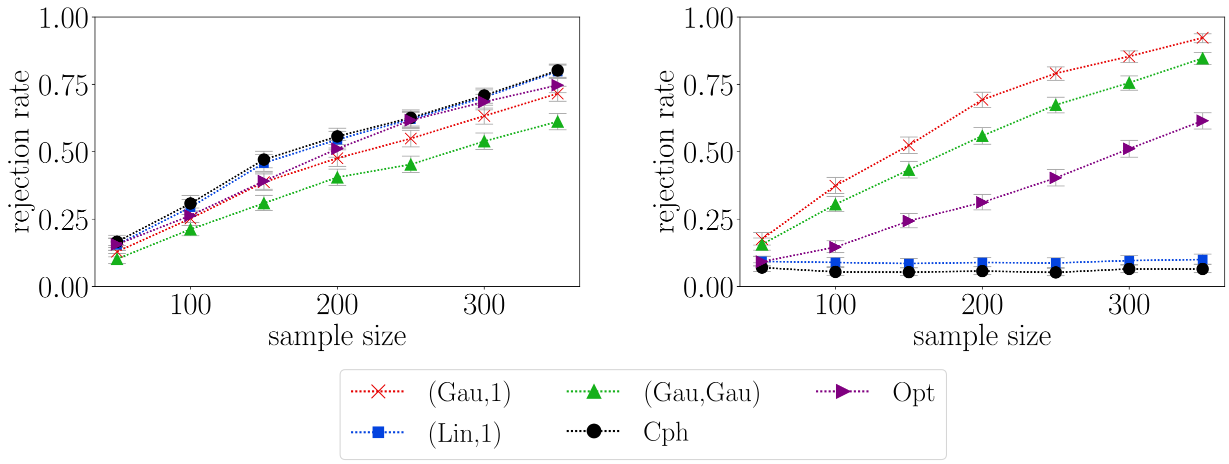

Figure 1: Left: the rejection rate of various tests for samples from Distribution 7. Right: the rejection rate of various tests for samples from Distribution 8.

Note that D.7 satisfies the Cox-proportional hazards assumption and is thus ideally suited for the Cph likelihood ratio test. Rejection rates are plotted in Figure 1 (left). The kernel log-rank test with the linear kernel on the covariates performs as well as the Cph test. This is to be expected because it models the hazard rate in the same way as the Cph test. While the Gaussian kernel does not make this assumption, it does not lose much power compared to the Cph test. The same holds for the (Gau, Gau) case. As expected, there is a trade-off between the richness of the kernel and the power of the test. The optHSIC method performs similar to the Gaussian kernel.

D.8: Nonlinear log-hazard: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{x^2\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 2.25)$.

Note that D.8 does not satisfy the CPH assumption. The rejection rate in Figure 1 (right) shows that the CPH and the kernel log-rank test with linear kernel are both unable to detect this dependence. By contrast, the (Gau,1) and (Gau,Gau) kernels are both able to detect the dependency and we note again a trade-off between richness and power. Both our tests outperform the optHSIC method.

In the next few sections we consider distributions in which the hazard function $d\Lambda_{Z|X}$ does not factorize into a product of a function of Z and a function of X.

D.9: Weibull distributions with different shapes: $Z|X = x \sim \text{Weib}(\text{shape} = 3.35 + 1.75 \cdot x, \text{scale} = 1)$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.75)$.

D.10: Normal distributions with different variances: $Z|X = x \sim \text{Normal}(\text{mean} = 100 - 2.25 \cdot x, \text{var} = 5.5 + x \cdot 4.5)$ and $C|X = x \sim 82 + \text{Exp}(\text{mean} = 35)$.

D.11: A checkerboard pattern: See Figure 2. Because the pattern is more complicated, we let the sample size range from 500 to 2000 in steps of 500.

Figure 5 in Section A.2 shows that the effect of time on the hazard rate differs widely per covariate in D.9-11. As a result, we need a kernel on both the times and the covariates. This is confirmed in the observed rejection rates presented in the right half of Figure 2. These show that the Cph likelihood ratio test and RKHS based tests with a constant kernel on the times do not detect the dependencies. The (Gau, Gau) kernel enables modelling of more complicated hazard functions, and detects dependencies even in these challenging cases. The optHSIC test has power too, but for D.9 and D.11 it has much less power than the kernel log-rank test.

Power for multidimensional covariates Let $X \sim \text{Normal}(\text{mean} = 0_d \text{ and cov} = \Sigma_d)$ where $0_d = (0, \dots, 0) \in \mathbb{R}_d$, $\Sigma_d = MM^T$ and M is a $d \times d$ matrix of independent $\text{Normal}(0, 1)$ entries. We study four distributions:

D.12: Cph dependence on all covariates: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{1_d^T x/20\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$ where $1_d = (1, \dots, 1) \in \mathbb{R}^d$.

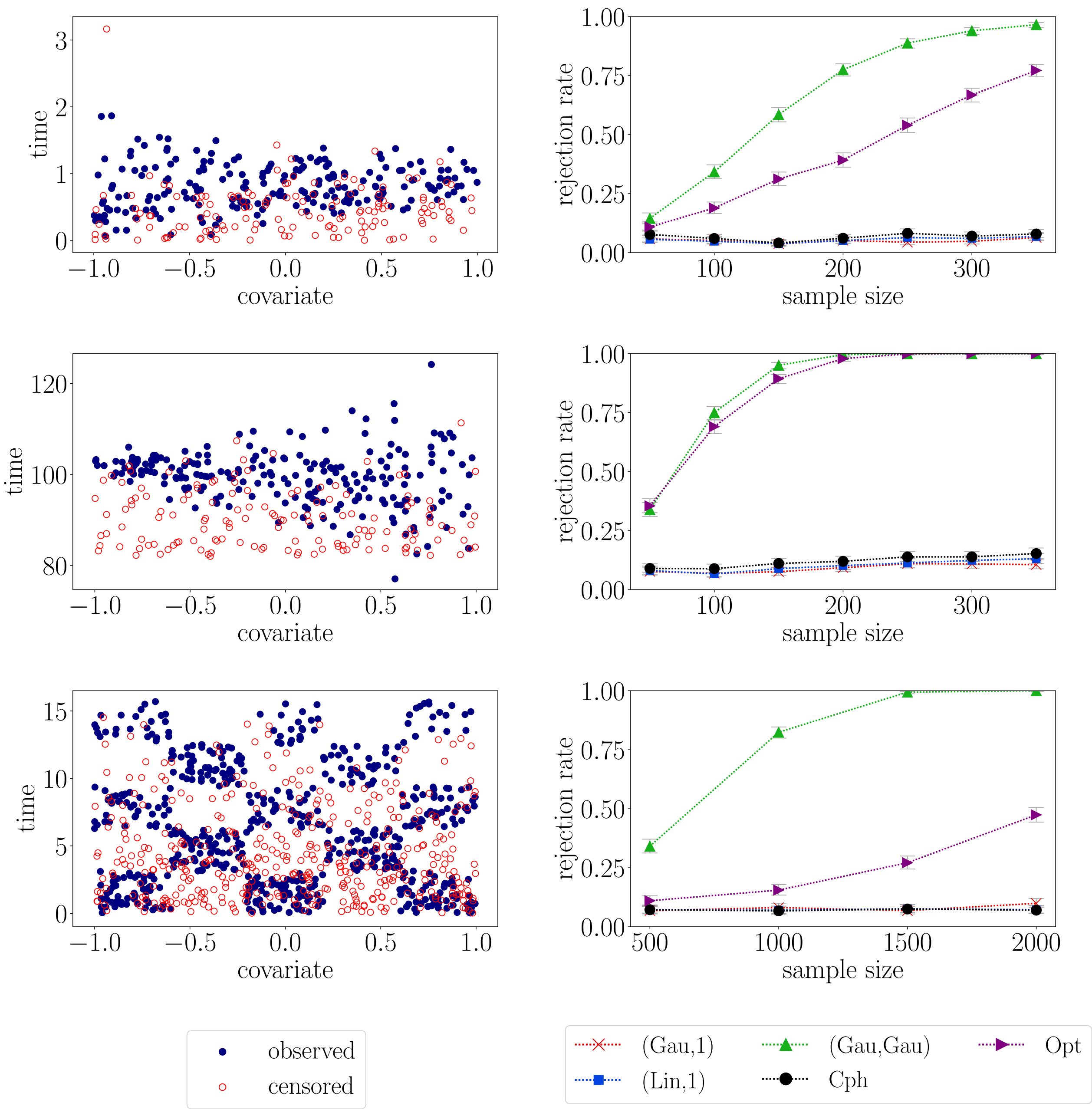

Figure 2: Left: From top to bottom, scatter-plots of samples from D.9,10 and 11. Right: corresponding rejection rates for each of the tests.

D.13: Cph dependence in a 1-dimensional subspace: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{x_1/60\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$.

D.14: Non-Cph dependence in a 1-dimensional subspace: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{x_1^2/60\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$.

D.15: Mixed dependence in a 2-dimensional subspace: $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{(x_1^2 + 3x_2)/60\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 2)$.

We investigate power both as n varies and as the dimension increases. Rejection rates are plotted in Figure 3. For distributions D.12 and D.13, all tests including the CPH likelihood ratio test perform roughly equally well. In D.14 the kernels $(\text{Gau}, 1)$ and (Gau, Gau) are the only ones to capture the non-CPH dependence. We note that optHSIC is outperformed by both $(\text{Gau}, 1)$ and (Gau, Gau). In D.15 there is a linear as well as a quadratic term present in the hazard, and the Gaussian kernel log-rank test performs best. As a final remark about multidimensional covariates, the Fisher kernel, like the CPH likelihood ratio test, standardizes the data, which is helpful in cases where centering and scaling of the data is not sufficient. An example is given in Section A.4 of the supplementary materials.

Varying censoring rates and distributions. We also studied the rejection rates in cases where censoring depends on the covariate and where the percentage of observed $(\Delta = 1)$ events is $15, 30, 45, 60, 75, 90$ or 100%. The distributions from which we sampled data and the obtained rejection rates are presented in Appendix A.5. Our main findings are that the proposed kernel log-rank test performs well for each censoring rate. In particular, the Type 1 error rate is correct for each of the censoring percentages. When the CPH assumption holds true, the kernel log-rank test still does not lose much power compared with the CPH likelihood ratio test. The trade-off between richness of the kernel and power remains similar to the presented in the main text. Moreover, the fact that richer kernels can be used to detect more complex relationships between covariates and survival times still holds true. A final important observation is that the optHSIC test has much lower power compared to the kernel log-rank test for high censoring percentages.

Sensitivity to choice of bandwidth. In the experiments presented thus far we set the bandwidth of the Gaussian kernel to be $\sigma^2 = \text{median}\{||x_i - x_j||^2 \ : i \neq j\}/2$. We now study the effect of varying the bandwidth of the Gaussian kernel on the rejection rate. The distributions used in the experiments, and tables containing the rejection rates for various bandwidths can be found in Section A.6. We find that, while for most scenarios a bandwidth can be selected that slightly outperforms the median heuristic, the median heuristic has a consistently good performance across all scenarios. We furthermore find that bandwidths much smaller than the median heuristic can lead to inflated Type 1 error rate.

For the kernel (Gau,Gau) we note that as the bandwidth of the times increases, the associated kernel matrix increasingly resembles an $n \times n$-matrix in which each entry equals 1. As a result, we find that the performance becomes more and more similar to the kernel (Gau,1), which we found could not detect some of the time-covariate interactions, but which had better power than the (Gau,Gau) kernel against scenarios that did not require a kernel on the times.

Choice of kernel Choosing a good kernel function is an important task when applying our testing procedure. While in principle we can choose $\mathfrak{K}$ as any kernel, choosing a kernel that factorizes into the products of two kernels, one for time and one for covariates, has the important advantage that it gives the simple closed-form expression for our test-statistic in Theorem 3.4. In Theorems 4.7 and 5.3, we prove that our testing procedure is consistent for sufficiently expressive kernels K and L.

Our experiments show that in some scenarios there is a trade off between richness of the RKHS and the power of the test. For example, when the data is sampled from a distribution satisfying the CPH assumption, the kernel $(\text{Lin}, 1)$ is generally more powerful than the richer $(\text{Gau}, 1)$ or (Gau, Gau) kernels. Hence, if we have prior knowledge about the relationship, one may use this knowledge to choose the appropriate kernels. For instance, if we know that there is a quadratic dependence on a 1-dimensional covariate, a suitable kernel would be $K(x, x') = x^2 x'^2$, which models this dependence. Or if we do not suspect time-covariate interactions to play an important role, we may choose the constant kernel on the times.

On the other hand, if we expect the dependency to be more complex, we may opt for a richer RKHS, such as the (Gau, Gau) kernel. We find this kernel has consistently good performance, and has competitive power also in the cases where less rich RKHSs are sufficient to detect the dependency.

One can also use a cross-validation approach to select both the kernel and possible parameters of the kernel (such as the bandwidth of the Gaussian kernel). That is, one can perform the Wild Bootstrap test on a training

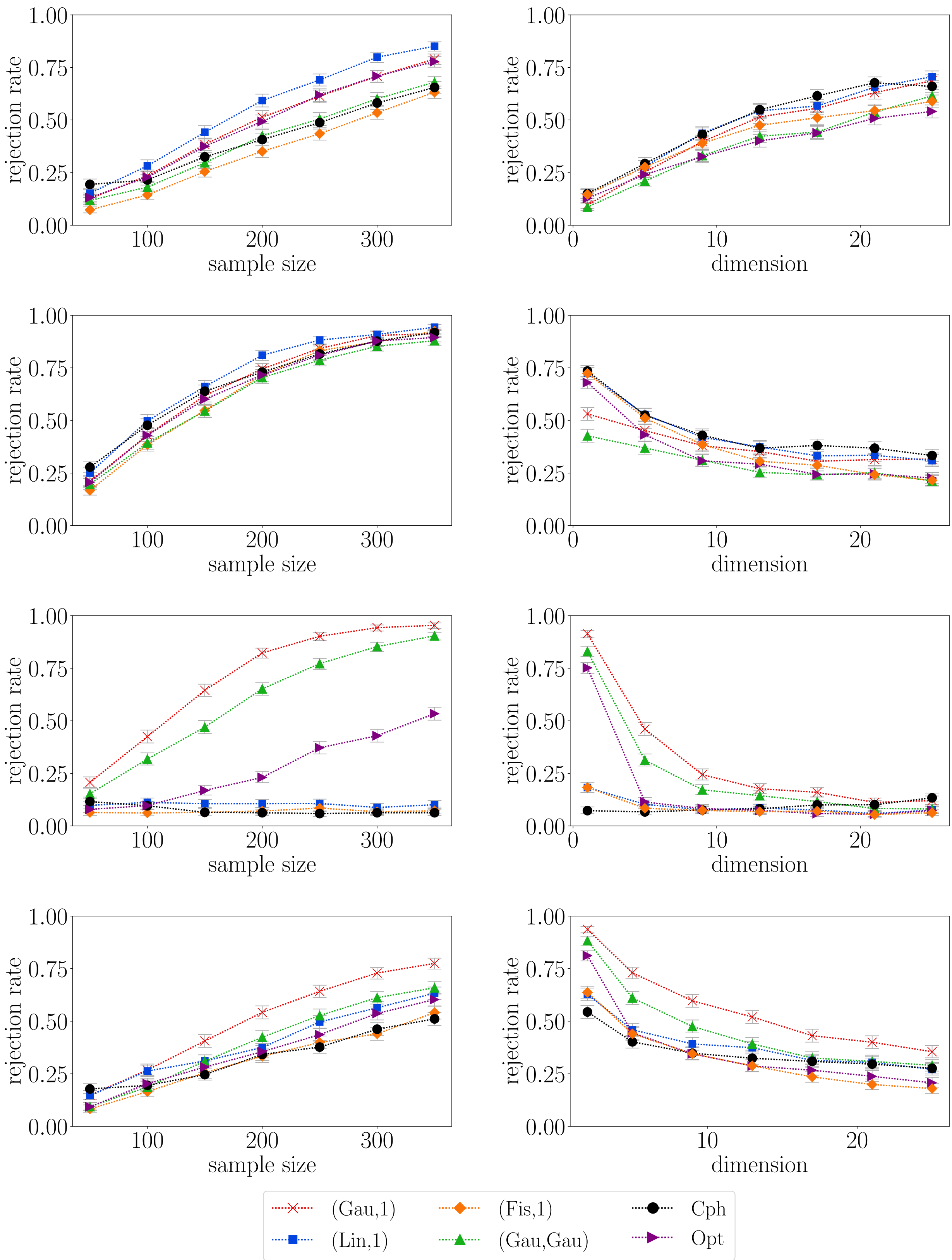

Figure 3: Left: rejection rates of the different tests against Distributions 12(top)-15(bottom) for $d = 10$ as the sample size varies. Right: rejection rates of the different tests against Distributions 16 (top) and 17 (bottom) as the dimension varies and $n = 200$.

subset of the data for different kernels and subsequently select the kernel (or parameter) that yielded the smallest p-value. The actual hypothesis test can then be done with the select kernel on the remaining part of the data. Also, a well-known approach in the uncensored setting is to choose the kernel parameters that maximize the power of the test. In the uncensored setting, the test power is increased by maximizing the ratio of the test statistic to its standard deviation under the alternative hypothesis [19]. We leave this as future research as for this approach we would need to find the asymptotic distribution of our test statistic under the alternative hypothesis.

7 Conclusions

We have introduced a novel non-parametric independence test between right-censored survival times Z and covariates X. Our approach uses an infinite-dimensional exponential family of cumulative hazard functions, which are parameterized by functions in a reproducing kernel Hilbert space. By choosing a expressive Hilbert space of functions, we show that our testing procedure is able to detect any type of dependence, while for very simple Hilbert spaces, we recover ubiquitous approaches such as the Cox-Score test. We provide a simple testing procedure based on Wild Bootstrap and show very good experimental results.

SUPPLEMENTARY MATERIAL

A. Additional experiments In Section A.1 we study the Type-I error, in Section A.2 we show additional figures, in Section A.3 we study the power for small deviations from the null-hypothesis, in Section A.4 we show some of the additional experimental results for high-dimensional covariates, in Section A.5 we show experiments for varying censoring percentages, and in Section A.6 we show experiments for varying bandwidths of the Gaussian kernel.

B. Preliminary results: In this section, and in order for this paper to be self-contained, we review some preliminary results that will be used on our proofs.

C. Auxiliary results: In this section we state some auxiliary results used by the proofs of the main results of the paper.

D. Main results: In this section we prove the main results of the paper.

E. Proofs of auxiliary results In this section we prove the auxiliary results introduced in Section C.

8 Acknowledgements

We thank David Steinsaltz for the helpful discussions early in the project. Also, we thank Nicolás Rivera for the helpful discussions and suggestions for this paper. Tamara Fernandez was supported by Biometrika Trust.

References

[1] Odd Aalen. Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.*, 6(3):534–545, 1978.

[2] Vilijandas Bagdonavičius, Julius Kruopis, and Mikhail S. Nikulin. *Non-parametric tests for censored data.* ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2011.

[3] Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics.* Kluwer Academic Publishers, Boston, MA, 2004. With a preface by Persi Diaconis.

[4] K. Chwialkowski and A. Gretton. A kernel independence test for random processes. In *ICML'14: Proceedings of the 31st International Conference on International Conference on Machine Learning*, page II–1422–II–1430. Proceedings of Machine Learning Research, 2014.

[5] David R. Cox. Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, 34:187–220, 1972.

[6] Herold Dehling and Thomas Mikosch. Random quadratic forms and the bootstrap for U-statistics. *J. Multivariate Anal.*, 51(2):392–413, 1994.

[7] Bradley Efron and Iain M. Johnstone. Fisher's information in terms of the hazard rate. *Ann. Statist.*, 18(1):38–62, 1990.

[8] Tamara Fernández and Nicolás Rivera. Kaplan-Meier V- and U-statistics. *Electron. J. Stat.*, 14(1):1872–1916, 2020.

[9] Tamara Fernández and Nicolás Rivera. A reproducing kernel hilbert space log-rank test for the two-sample problem. *To appear: Scandinavian Journal of Statistics.*

[10] 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.

[11] Valérie Garès, Sandrine Andrieu, Jean-François Dupuy, and Nicolas Savy. An omnibus test for several hazard alternatives in prevention randomized controlled clinical trials. *Stat. Med.*, 34(4):541–557, 2015.

[12] R. D. Gill. *Censoring and stochastic integrals*, volume 124 of *Mathematical Centre Tracts.* Mathematisch Centrum, Amsterdam, 1980.

[13] Richard Gill. Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, 11(1):49–58, 1983.

[14] Robert J. Gray. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87(420):942–951, 1992.

[15] A. Gretton, O. Bousquet, A. J. Smola, and B. Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In *Algorithmic Learning Theory: 16th International Conference*, pages 63–77. Springer-Verlag, 2005.

[16] A. Gretton, K. Fukumizu, C.-H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. In *Advances in Neural Information Processing Systems 20*, pages 585–592. MIT Press, 2008.

[17] Arthur Gretton. A simpler condition for consistency of a kernel independence test. *arXiv preprint arXiv:1501.06103*, 2015.

[18] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012.

[19] Arthur Gretton, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25, pages 1205–1213. Curran Associates, Inc., 2012.

[20] David P. Harrington and Thomas R. Fleming. A class of rank test procedures for censored survival data. *Biometrika*, 69(3):553–566, 1982.

[21] Wolfgang Kössler. Max-type rank tests, U-tests, and adaptive tests for the two-sample location problem—an asymptotic power study. *Comput. Statist. Data Anal.*, 54(9):2053–2065, 2010.

[22] Chap T. Le, Patricia M. Grambsch, and Thomas A. Louis. Association between survival time and ordinal covariates. *Biometrics*, 50(1):213–219, 1994.

[23] Nathan Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, 50(3):163, 1966.

[24] Ian W. McKeague, A. M. Nikabadze, and Yan Qing Sun. An omnibus test for independence of a survival time from a covariate. *Ann. Statist.*, 23(2):450–475, 1995.

[25] Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.

[26] 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.

[27] David Rindt, Dino Sejdinovic, and David Steinsaltz. Nonparametric independence testing for right-censored data using optimal transport. *arXiv preprint arXiv:1906.03866*, 2019.

[28] Robert J. Serfling. *Approximation theorems of mathematical statistics.* John Wiley & Sons, Inc., New York, 1980. Wiley Series in Probability and Mathematical Statistics.

[29] Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In Marcus Hutter, Rocco A. Servedio, and Eiji Takimoto, editors, *Algorithmic Learning Theory*, pages 13–31, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[30] Bharath K. Sriperumbudur, Kenji Fukumizu, and Gert R. G. Lanckriet. Universality, characteristic kernels and RKHS embedding of measures. *J. Mach. Learn. Res.*, 12:2389–2410, 2011.

[31] Robert E. Tarone. On the distribution of the maximum of the logrank statistic and the modified wilcoxon statistic. *Biometrics*, 37(1):79–85, 1981.

[32] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[33] Qinyi Zhang, Sarah Filippi, Arthur Gretton, and Dino Sejdinovic. Large-scale kernel methods for independence testing. *Stat. Comput.*, 28(1):113–130, 2018.

[34] David M. Zucker and Alan F. Karr. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.*, 18(1):329–353, 1990.

A Additional experiments

Here we present some experiments that were not described in detail in the main text.

A.1 Section 6: Distributions and tables used to study Type I error rate

To estimate the false rejection rate of our proposed test, we repeatedly take samples from distributions in which $Z \perp\!\!\!\perp X$ and we count the portion of experiments in which the kernel log-rank test rejects the null hypothesis. We set the significance level at $\alpha = 0.05$ and let the sample sizes range from 50 to 350 in steps of 50 and for each distribution and sample size, we take 5000 samples. Table 1 shows the distributions we sample from. Figure 4 shows scatterplots of samples from distributions 1-4.

D.	$Z\|X$	$C\|X$	X
1	Exp(mean $= 0.66$)	Exp(mean $= \exp(X/3)$)	Unif[-1,1]
2	Exp(mean $= 0.9$)	Exp(mean $= \exp(X^2)$)	Unif[-1,1]
3	Exp(mean $= 0.9$)	Weib(mean $= 3.25 + 1.75X$)	Unif[-1,1]
4	Exp(mean $= 0.9$)	$1 + X$	Unif[-1,1]
5	Exp(mean $= 0.6$)	Exp(mean $= \exp(1^T X)$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$
6	Exp(mean $= 0.6$)	Exp(mean $= \exp(X_1)$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$

Table 1: The distributions used to restimate the type 1 error rate. Exp(mean) $= \mu$ denotes the exponential distribution with mean μ. We define $\Sigma_{10} = MM^T$ where M is a 10×10 matrix of i.i.d. standard normal entries. Parameters are chosen such that approximately 60% of events is observed ($\Delta = 1$).

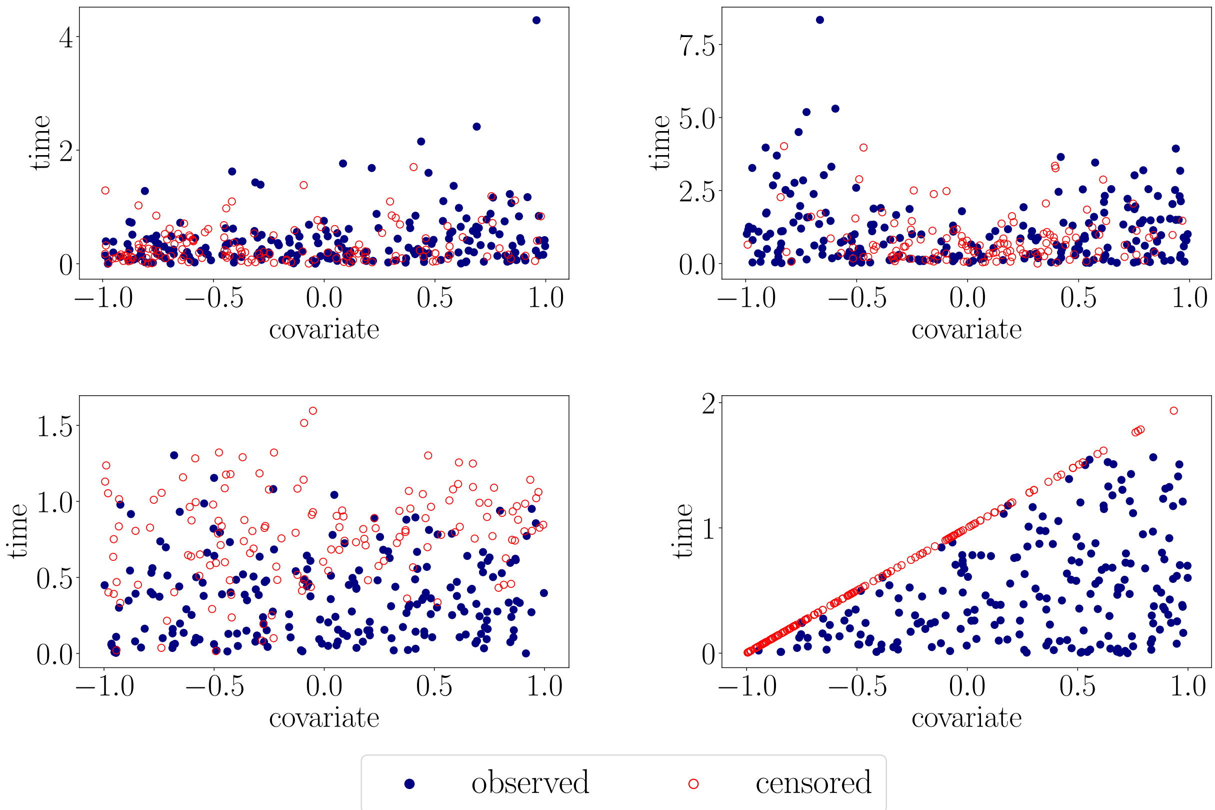

Figure 4: Top: Scatterplots of samples of size 350 from Distributions 1 (left) and 2 (right) of Table 1. Bottom: Scatterplots of samples of size 350 from Distributions 3 (left) and 4 (right).

We estimate the bandwidth of the Gaussian kernel from the same data as we apply the test to. Similarly we estimate the Fisher information matrix on the same data as we apply the kernel log-rank test to. Due to this dependence of the kernel on the data, the arguments of Section 5 do not apply directly. The obtained rejection rates show that this does not lead to increased type 1 error rate. The resulting rejection rates are displayed in Tables 2 and 3.

D.	Method	Sample size $n =$ 50	100	150	200	250	300	350
1	Cph	0.053	0.051	0.049	0.051	0.055	0.048	0.053
	(Lin,1)	0.037	0.049	0.050	0.047	0.055	0.048	0.044
	(Gau,1)	0.046	0.046	0.045	0.050	0.045	0.052	0.051
	(Gau,Gau)	0.047	0.044	0.044	0.048	0.045	0.054	0.052
	Opt	0.050	0.051	0.055	0.046	0.051	0.053	0.054
2	Cph	0.049	0.060	0.051	0.049	0.056	0.046	0.050
	(Lin,1)	0.044	0.044	0.050	0.047	0.052	0.045	0.047
	(Gau,1)	0.038	0.044	0.048	0.047	0.052	0.057	0.049
	(Gau,Gau)	0.043	0.045	0.051	0.049	0.047	0.050	0.051
	Opt	0.053	0.050	0.051	0.052	0.047	0.050	0.054
3	Cph	0.045	0.052	0.052	0.052	0.049	0.052	0.056
	(Lin,1)	0.039	0.047	0.048	0.048	0.053	0.046	0.046
	(Gau,1)	0.039	0.047	0.041	0.052	0.052	0.051	0.051
	(Gau,Gau)	0.048	0.046	0.051	0.049	0.051	0.052	0.049
	Opt	0.057	0.051	0.055	0.049	0.046	0.053	0.050
4	Cph	0.050	0.051	0.051	0.049	0.052	0.052	0.047
	(Lin,1)	0.047	0.049	0.052	0.049	0.049	0.051	0.052
	(Gau,1)	0.045	0.046	0.048	0.045	0.044	0.053	0.047
	(Gau,Gau)	0.043	0.051	0.049	0.045	0.049	0.052	0.048
	Opt	0.050	0.050	0.046	0.046	0.046	0.047	0.052

Table 2: The type 1 error rates of the various methods under Distributions 1-4, which are found in Table 1. The covariates are 1-dimensional. Rejection rates above 0.057 are displayed in italics.

D.	Method	Sample size $n =$ 50	100	150	200	250	300	350
5	Cph	*0.122*	*0.080*	*0.064*	*0.068*	*0.067*	*0.060*	*0.060*
	(Lin,1)	0.050	0.051	0.053	0.056	0.050	0.054	*0.058*
	(Fis,1)	0.054	0.045	0.051	0.051	0.050	0.048	0.045
	(Gau,1)	0.046	0.041	0.047	0.053	0.053	0.054	0.054
	(Gau,Gau)	0.044	0.043	0.040	0.049	0.053	0.048	0.049
	Opt	0.052	0.054	0.054	0.047	0.049	0.052	0.054
6	Cph	*0.123*	*0.076*	*0.065*	*0.065*	*0.060*	*0.061*	0.056
	(Lin,1)	0.054	0.051	0.051	0.053	0.046	0.054	*0.058*
	(Fis,1)	0.050	0.048	0.054	0.050	0.050	0.049	0.047
	(Gau,1)	0.044	0.046	0.049	0.051	0.049	0.054	0.049
	(Gau,Gau)	0.047	0.045	0.044	0.053	0.048	0.045	0.054
	Opt	0.049	0.049	0.053	0.050	0.054	0.056	0.047

Table 3: The type 1 error rates of the various methods under Distributions 5-6, which are found in Table 1. The covariates are 10-dimensional. Rejection rates above 0.057 are in italics.

A.2 Section 6: Hazard rates of distributions with time-covariate interactions

Recall D.9-11 from Section 6. These were:

D.9: **Weibull distributions with different shapes:** $Z|X = x \sim \text{Weib}(\text{shape} = 3.35 + 1.75 \cdot x, \text{scale} = 1)$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.75)$

D.10: **Normal distributions with different variances:** $Z|X = x \sim \text{Normal}(\text{mean} = 100 - 2.25 \cdot x, \text{var} = 5.5 + x \cdot 4.5)$ and $C|X = x \sim 82 + \text{Exp}(\text{mean} = 35)$

D.11: **A checkerboard pattern:** See Figure 2.

The hazard rates of these three distributions are given in Figure 5.

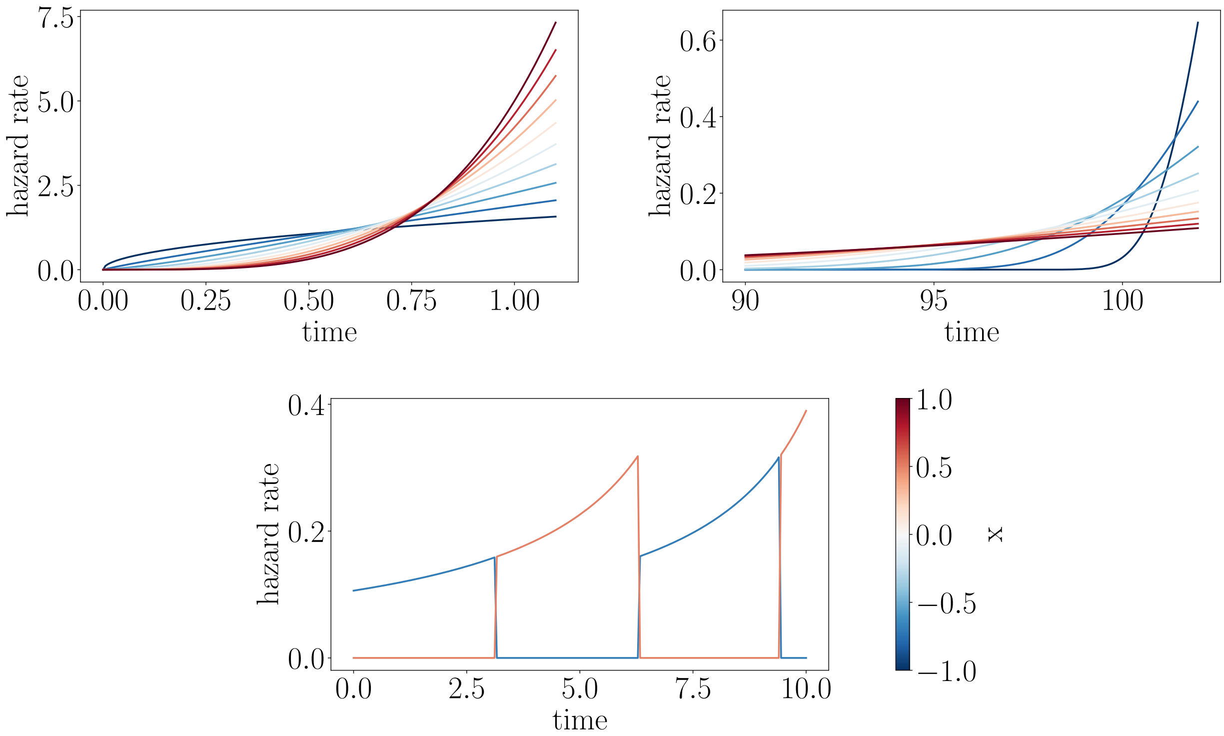

Figure 5: Top row left: the hazard rate as a function of time for the Weibull distribution of distribution D.9. Different lines correspond to different values of x. Top row right: the hazard rate for the normal distribution of distribution D.10. Bottom row: the hazard rate of the checkerboard pattern of distribution D.11.

A.3 Simulations of deviations from the null-hypothesis

In the main text we investigated the power performance at different sample sizes. We now fix the sample size to 100 and study the power of the various tests for small deviations from the null hypothesis. For this, we sample $X \sim \text{Normal}(\mu = 0, \sigma^2 = 1)$ and consider two different settings in which θ varies from -0.4 to 0.4 in steps of 0.1. Notice that, for the two settings described below, $\theta = 0$ recovers the null hypothesis.

1. **Cph distributions:** $T|X = x \sim \text{Exp}(\text{mean} = \exp\{\theta \cdot x\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$
2. **Non-linear log-hazard distributions:** $T|X = x \sim \text{Exp}(\text{mean} = \exp\{\theta \cdot x^2\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$

The plots in Figure 6 below show the rejection rates of each method.

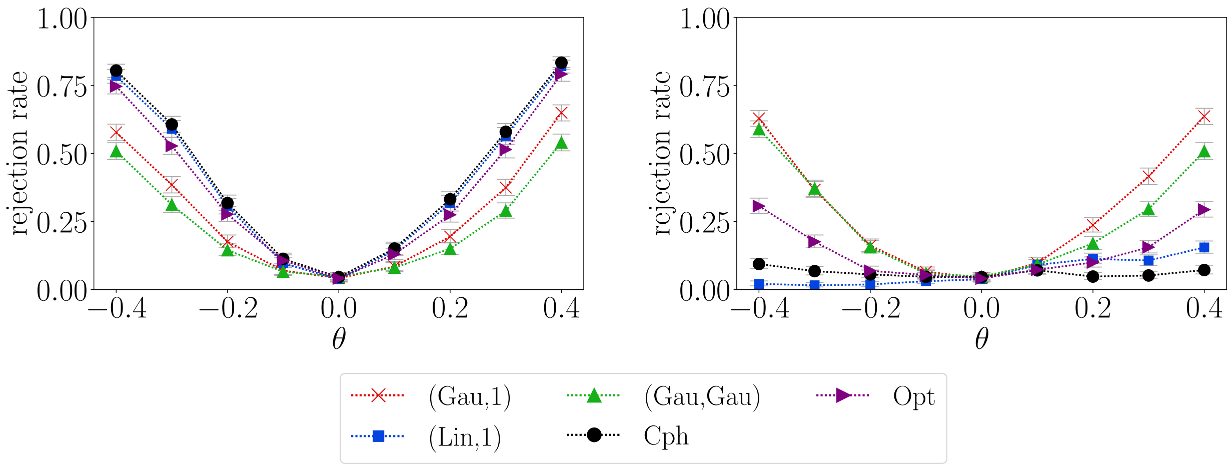

Figure 6: Left: the rejection rate of the various tests against the Cph distributions. Right: the rejection rate of the various tests against the non-linear log-hazard distributions.

A.4 Section 6: Power for higher dimensional covariates

We now provide an example of the usefulness of the Fisher-kernel. When assessing the power in high-dimensional scenarios all the kernels $(\text{Lin}, 1)$, $(\text{Gau}, 1)$ and (Gau, Gau) perform consistently well, thus the usefulness of the Fisher kernel is not obvious. The Fisher kernel has the ability to 'standardize' the data in cases where simple centering and scaling of the individual covariate dimensions is not sufficient. To illustrate this, we consider the following example. Let $\Sigma_{11} = 1/10$, $\Sigma_{ii} = 1$ for $i > 1$ and $\Sigma_{ij} = 0$ otherwise, and let R be an orthogonal rotation matrix. Let $X \sim \text{Normal}(0, R\Sigma R^T)$ and let $v = (1, 0, ..., 0)^T$.

D.16: **A distribution in which the Fisher information helps uncover the dependency:** Let $Z|X = x \sim \text{Exp}(\text{mean} = \exp\{v^T(R^T x)\})$ and $C|X = x \sim \text{Exp}(\text{mean} = 1.5)$

Figure 7 shows that the RKHS test with kernels (Lin,1) and (Gau,1) does not detect the dependency. The Fisher and Cph likelihood ratio tests however have power, because the inverse information matrix has the effect of standardizing the data.

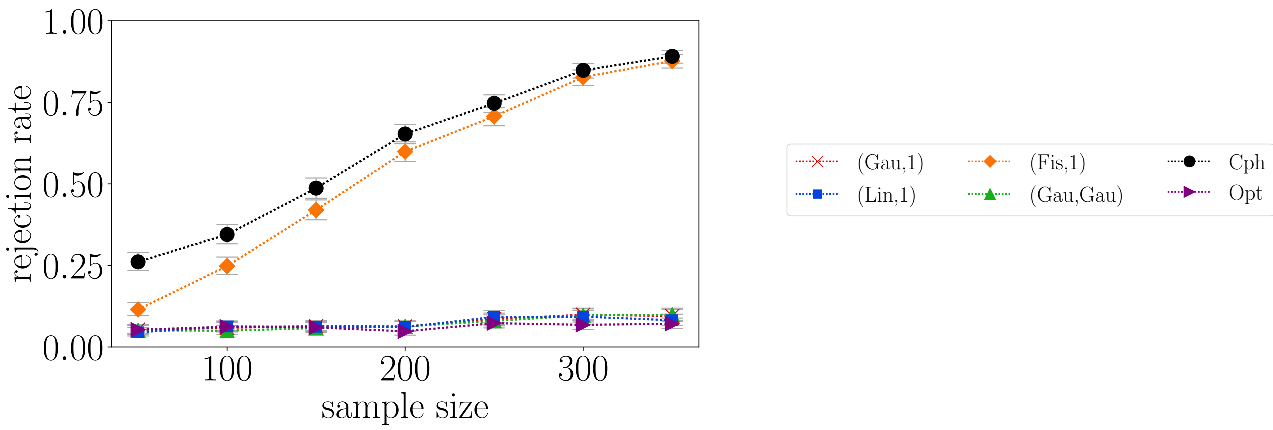

Figure 7: The rejection rate of the different methods for distribution D.17, showing the effect of the Fisher information matrix.

A.5 Dependent censoring and varying censoring rates

We now test how the wild bootstrap test using the kernel log-rank statistic performs under varying censoring distributions. In particular we parametrize the censoring distributions and vary the parameters to find estimate the power of each method when 15, 30, 45, 60, 75, 90 or 100% of the observations is uncensored. We compare power and type 1 error to methods such as optHSIC (Opt) and Cox-proportional-hazards Likelihood ratio test (Cph). We study the type 1 error rate in Section A.5.1 and the power in Section A.5.2.

A.5.1 Type 1 error under varying censoring rates

The distributions from which we sample data in which $X \perp\!\!\!\perp T$ are given in Table 4. In each case the sample size $n = 200$. To estimate the rejection rates we sampled 5000 times, and counted the number of times we rejected the null. Obtained rejection rates are displayed in Tables 5 and 6. In each of the scenarios we find the type 1 error rate to be correct for each censoring percentage and for each kernel.

D.	$Z\|X$	$C\|X$	X
17	Exp(mean = 1)	Exp(mean = θ)	Unif[-1,1]
18	Exp(mean = 2.2)	Exp(mean = $\theta\exp(X)$)	Unif[-1,1]
19	Weib(shape = 3.25)	Exp(mean = θX^2)	Unif[-1,1]
20	$\mathcal{N}(\text{mean} = 99, \text{var} = 5.5)$	Exp(mean = $82 + \theta$)	Unif[-1,1]
21	Exp(mean = 1)	Exp(mean = θ)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$
22	Exp(mean = 1)	Exp(mean = $\theta\exp(1^T X/30)$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$
23	Exp(mean = exp(0.5)	Exp(mean = $\theta\exp(X_2^2)/20$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$

Table 4: The parametrized distributions to test the type 1 error rate under different censoring rates. Here $\Sigma_{10} = MM^T$ where M is a 10×10 matrix of i.i.d. standard normal entries. The parameter θ is varies such that $15, 30, 45, 60, 75, 90, 100\%$ of the individuals are observed (i.e. $\Delta = 1$). The sample size is $n = 200$ in each case.

		% Observed						
D.	Method	15%	30%	45%	60%	75%	90%	100%
17	Cph	0.053	0.050	0.047	0.049	0.052	0.050	0.046
	(Lin,1)	0.051	0.046	0.046	0.046	0.049	0.048	0.044
	(Gau,1)	0.046	0.045	0.047	0.050	0.053	0.048	0.044
	(Gau,Gau)	0.045	0.046	0.044	0.046	0.056	0.053	0.045
	Opt	0.048	0.047	0.052	0.045	0.053	0.046	0.045
18	Cph	0.050	0.050	0.049	0.055	0.050	0.057	0.046
	(Lin,1)	0.049	0.049	0.047	0.053	0.050	0.055	0.046
	(Gau,1)	0.050	0.050	0.047	0.045	0.050	0.053	0.045
	(Gau,Gau)	0.046	0.048	0.044	0.044	0.048	0.052	0.049
	Opt	0.053	0.055	0.050	0.051	0.050	0.053	0.049
19	Cph	0.054	0.051	0.050	0.055	0.053	0.054	0.046
	(Lin,1)	0.044	0.048	0.047	0.052	0.051	0.054	0.046
	(Gau,1)	0.043	0.048	0.044	0.049	0.054	0.050	0.046
	(Gau,Gau)	0.042	0.047	0.040	0.048	0.053	0.050	0.047
	Opt	0.054	0.052	0.050	0.053	0.051	0.052	0.050
20	Cph	0.055	0.051	0.054	0.050	0.054	0.052	0.050
	(Lin,1)	0.053	0.047	0.049	0.047	0.051	0.050	0.049
	(Gau,1)	0.049	0.045	0.048	0.047	0.050	0.051	0.051
	(Gau,Gau)	0.044	0.047	0.046	0.044	0.052	0.050	0.048
	Opt	0.050	0.044	0.049	0.047	0.055	0.053	0.049

Table 5: The type 1 error rates of the various methods under different censoring rates. The Distributions 17-20 are found in Table 4. The covariates are 1-dimensional. There are no rejection rates larger than 0.057.

D.	Method	% Observed 15%	30%	45%	60%	75%	90%	100%
21	Cph	*0.063*	*0.061*	*0.068*	0.056	*0.062*	*0.058*	*0.062*
	(Lin,1)	0.046	0.043	0.056	0.047	0.048	0.048	0.049
	(Fis,1)	0.048	0.047	0.057	0.047	0.049	0.046	0.050
	(Gau,1)	0.044	0.044	0.056	0.046	0.046	0.048	0.050
	(Gau,Gau)	0.040	0.047	0.053	0.048	0.046	0.046	0.049
	Opt	0.053	0.045	*0.059*	0.048	0.048	0.049	0.054
22	Cph	*0.061*	*0.066*	*0.066*	*0.064*	*0.063*	*0.066*	*0.063*
	(Lin,1)	0.051	0.054	0.056	0.048	0.050	0.052	0.050
	(Fis,1)	0.049	0.050	0.051	0.049	0.053	0.052	0.051
	(Gau,1)	0.048	0.049	0.055	0.050	0.052	0.052	0.052
	(Gau,Gau)	0.050	0.048	0.057	0.048	0.050	0.048	0.053
	Opt	0.055	0.051	*0.058*	0.049	0.048	0.052	0.053
23	Cph	*0.068*	*0.068*	*0.067*	*0.064*	*0.064*	*0.064*	*0.063*
	(Lin,1)	0.048	0.054	0.055	0.048	0.049	0.050	0.050
	(Fis,1)	0.051	0.055	0.055	0.046	0.050	0.053	0.051
	(Gau,1)	0.047	0.049	0.053	0.047	0.050	0.050	0.052
	(Gau,Gau)	0.047	0.048	*0.059*	0.048	0.051	0.049	0.053
	Opt	0.047	0.049	0.053	0.054	0.051	0.051	0.053

Table 6: The type 1 error rate of the various methods under different censoring rates. The Distributions 21-23 are found in Table 4. The covariates are 10-dimensional. Rejection rates above 0.057 are displayed in italics.

A.5.2 Power under varying censoring rates

We now estimate the power when the alternative hypothesis holds. To estimate power we sample 1000 times from each distribution of Table 7 and count the number of rejections. In each case the sample size $n = 200$. Obtained rejection rates are displayed in Tables 8 and 9.

D.	$Z\|X$	$C\|X$	X
24	Exp(mean $= \exp(X/3)$)	Exp(mean $= \theta$)	Unif[-1,1]
25	Exp(mean $= \exp(X^2)$)	Exp(mean $= \theta \exp(X)$)	Unif[-1,1]
26	Weib(shape $= 1.5X + 3.25$)	Exp(mean $= \theta X^2$)	Unif[-1,1]
27	$\mathcal{N}$(mean $= 100 - X$, var $= 1.5X + 5.5$)	Exp(mean $= 82 + \theta$)	Unif[-1,1]
28	Exp(mean $= \exp(1^T X/30)$)	Exp(mean $= \theta$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$
29	Exp(mean $= \exp(X_4/7)$)	Exp(mean $= \theta \exp(1^T X/30)$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$
30	Exp(mean $= \exp(X_4^2/20)$)	Exp(mean $= \theta \exp(X_2^2)/20$)	$\mathcal{N}_{10}(0, \text{cov} = \Sigma_{10})$

Table 7: The parametrized distributions to test the power under different censoring rates. Here $\Sigma_{10} = MM^T$ where M is a 10×10 matrix of i.i.d. standard normal entries. The parameter θ is varies such that $15, 30, 45, 60, 75, 90, 100\%$ of the individuals are observed (i.e. $\Delta = 1$). The sample size is $n = 200$ in each case.

		% Observed						
D.	Method	15%	30%	45%	60%	75%	90%	100%
24	Cph	**0.16**	**0.30**	**0.44**	**0.56**	**0.67**	**0.72**	**0.76**
	(Lin,1)	**0.16**	0.28	**0.44**	0.55	0.66	0.71	**0.76**
	(Gau,1)	0.14	0.24	0.36	0.48	0.58	0.64	0.68
	(Gau,Gau)	0.12	0.20	0.31	0.40	0.48	0.54	0.57
	Opt	**0.16**	0.29	0.41	0.52	0.62	0.68	0.71
25	Cph	0.06	0.10	0.10	0.10	0.08	0.08	0.06
	(Lin,1)	0.09	0.13	0.14	0.14	0.12	0.11	0.09
	(Gau,1)	**0.22**	**0.41**	**0.56**	**0.68**	**0.79**	**0.85**	**0.90**
	(Gau,Gau)	0.16	0.31	0.42	0.54	0.66	0.77	0.84
	Opt	0.08	0.13	0.19	0.26	0.40	0.52	0.63
26	Cph	0.11	0.06	0.08	0.12	0.17	0.21	0.24
	(Lin,1)	0.10	0.05	0.06	0.09	0.14	0.19	0.22
	(Gau,1)	0.10	0.04	0.05	0.08	0.12	0.16	0.16
	(Gau,Gau)	**0.38**	**0.54**	**0.68**	**0.76**	**0.84**	0.88	**0.87**
	Opt	0.22	0.16	0.18	0.34	0.63	**0.89**	0.86
27	Cph	0.29	0.12	0.08	0.05	0.06	0.12	0.14
	(Lin,1)	0.27	0.11	0.06	0.04	0.06	0.10	0.12
	(Gau,1)	0.22	0.09	0.07	0.04	0.05	0.09	0.11
	(Gau,Gau)	0.42	**0.51**	**0.62**	**0.76**	**0.86**	0.94	0.95
	Opt	**0.54**	0.42	0.39	0.57	0.85	**0.98**	**0.99**

Table 8: The power of the various methods under different censoring rates. The Distributions 24-27 are found in Table 7. The covariates are 1-dimensional. The highest rejection rate for each scenario and percentage of observed events is in bold. If an elevated type 1 error rate was found for a method under the censoring distribution in question - see Table 5, then the rejection rate is displayed in italics.

		% Observed						
D.	Method	15%	30%	45%	60%	75%	90%	100%
28	Cph	***0.22***	*0.40*	***0.58***	*0.68*	*0.83*	*0.88*	*0.92*
	(Lin,1)	0.21	**0.44**	**0.58**	**0.70**	**0.84**	**0.90**	**0.94**
	(Fis,1)	0.17	0.34	0.53	0.64	0.79	0.86	0.90
	(Gau,1)	0.20	0.40	0.55	0.67	0.82	0.88	0.91
	(Gau,Gau)	0.16	0.30	0.46	0.56	0.73	0.78	0.83
	Opt	0.18	0.33	0.49	0.59	0.74	0.82	0.91
29	Cph	*0.34*	*0.60*	*0.81*	*0.92*	*0.96*	*0.98*	***1.00***
	(Lin,1)	**0.39**	**0.64**	**0.87**	**0.95**	**0.98**	**0.99**	**1.00**
	(Fis,1)	0.29	0.54	0.76	0.90	0.96	0.97	0.99
	(Gau,1)	0.33	0.59	0.81	0.92	0.97	0.97	0.99
	(Gau,Gau)	0.24	0.49	0.73	0.84	0.93	0.95	0.98
	Opt	0.30	0.53	0.74	0.86	0.95	0.96	0.99
30	Cph	*0.06*	*0.08*	*0.07*	*0.08*	*0.09*	*0.08*	*0.11*
	(Lin,1)	0.06	0.06	0.08	0.10	0.15	0.15	0.18
	(Fis,1)	0.04	0.07	0.06	0.07	0.08	0.08	0.10
	(Gau,1)	**0.08**	**0.14**	**0.28**	**0.48**	**0.76**	**0.91**	**0.96**
	(Gau,Gau)	0.07	0.11	0.18	0.29	0.52	0.65	0.75
	Opt	0.05	0.06	0.08	0.14	0.25	0.40	0.81

Table 9: The power of the various methods under different censoring rates. The Distributions 28-30 are found in Table 7. The covariates are 10-dimensional. The highest rejection rate for each scenario and percentage of observed events is in bold. If an elevated type 1 error rate was found for a method under the censoring distribution in question (see Table 6), then the rejection rate is displayed in italics.

A.6 Varying bandwidths of the Gaussian Kernel

We now study how the rejection rates of the kernel log-rank test with kernel $(\mathrm{Gau}, 1)$ or $(\mathrm{Gau}, \mathrm{Gau})$ is affected by the choice of the bandwidth. Note that data is scaled so that each component has mean zero and unit variance. In this section $n = 200$ and 60 or 75% of the observations is observed. In Sections A.6.1 and A.6.2 we study the type 1 error rate and power of $(\mathrm{Gau}, 1)$ respectively, and in Sections A.6.3 and A.6.4 we study the type 1 error rate and power of $(\mathrm{Gau}, \mathrm{Gau})$ respectively. For estimation of type 1 error rate we sample 5000 times from each distribution, and for estimation of power we sample 1000 times from each distribution.

A.6.1 Type 1 error rate of the kernel (Gau,1) for different bandwidths

Obtained rejection rates are displayed in Table 10. We find that when the bandwidth is too small relative to the median heuristic, the type 1 error rate is inflated.

		$\sigma_X =$								
D.	$\Delta = 1$	σ_{med}	0.1	0.2	0.5	1	2	5	10	20
17	60%	0.049	0.044	0.044	0.051	0.050	0.047	0.050	0.048	0.045
18	60%	0.050	0.047	0.045	0.047	0.053	0.046	0.055	0.050	0.047
19	75%	0.050	0.051	0.049	0.043	0.049	0.046	0.056	0.048	0.053
20	75%	0.050	0.047	0.050	0.046	0.053	0.053	0.048	0.053	0.042
21	75%	0.048	*0.068*	*0.108*	0.049	0.043	0.050	0.050	0.046	0.048
22	60%	0.049	*0.079*	*0.154*	0.054	0.048	0.045	0.052	0.056	0.052
23	75%	0.046	*0.075*	*0.127*	0.046	0.043	0.052	0.050	0.051	0.050

Table 10: The type 1 error rate of the kernel log-rank test with kernel $(\mathrm{Gau}, 1)$ with various bandwidths. Distributions can be found in Table 5. For distributions D.17-20 it holds that $\sigma_{\mathrm{med}} = 0.72$ and for D.21-23 it holds that $\sigma_{\mathrm{med}} = 2.96$. The third column lists the rejection rate when the sigma median heuristic is used as bandwidth. The remaining columns list rejection rates when the bandwidths $0.1 - 20$ are used. Rejection rates above 0.57 are displayed in italics.

A.6.2 Power of the kernel (Gau,1) for different bandwidths

		$\sigma_X =$								
D.	$\Delta = 1$	σ_{med}	0.1	0.2	0.5	1	2	5	10	20
24	60%	0.48	0.21	0.32	0.44	0.51	**0.54**	0.53	**0.54**	0.53
25	60%	0.68	0.50	0.62	**0.69**	0.60	0.29	0.15	0.14	0.15
26	75%	0.12	0.10	0.11	0.11	0.13	0.14	0.14	**0.16**	0.14
27	75%	0.05	**0.07**	0.06	0.07	0.05	0.06	0.05	0.05	0.05
28	75%	0.59	*0.09*	*0.17*	0.07	0.27	0.52	**0.66**	0.61	0.63
29	60%	0.85	*0.09*	*0.17*	0.09	0.42	0.79	0.86	**0.89**	0.87
30	75%	0.67	*0.03*	*0.07*	0.09	0.59	**0.75**	0.40	0.19	0.13

Table 11: The power of the kernel log-rank test with kernel $(\mathrm{Gau}, 1)$ with various bandwidths. Distributions are found in Table 7. For distributions D.17-20 it holds that $\sigma_{\mathrm{med}} = 0.72$ and for D.21-23 it holds that $\sigma_{\mathrm{med}} = 2.96$. The third column lists the rejection rate when the sigma median heuristic is used as bandwidth. The remaining columns list rejection rates when the bandwidths $0.1 - 20$ are used. The highest rejection rate is in boldface. If the associated censoring distribution a bandwidth led to an elevated type 1 error rate (above 0.57) in Section A.6.1, then the rejection rate is displayed in italics.

A.6.3 Type 1 error rate of the kernel (Gau,Gau) for different bandwidths

		$\sigma_X =$							
		0.72	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.16	0.050	0.044	0.050	0.049	0.050	0.043	0.045	0.046
	0.10	0.051	0.045	0.050	0.051	0.045	0.044	0.049	0.052
	0.20	0.050	0.045	0.052	0.050	0.051	0.051	0.054	0.050
	0.50	0.045	0.049	0.053	0.051	0.052	0.044	0.045	0.047
	1.00	0.048	0.052	0.047	0.052	0.045	0.048	0.051	0.054
	2.00	0.045	0.045	0.046	0.048	0.052	0.049	0.046	0.052
	5.00	0.051	0.043	0.044	0.049	0.050	0.047	0.051	0.050
	10.00	0.046	0.049	0.049	0.046	0.049	0.047	0.050	0.045

Table 12: Type 1 error rate of the kernel log-rank test with various combinations of bandwidths for $D.19$ of Table 4 when 75% of events is observed. The median heuristic yields $\sigma_X = 0.72$ (first column) and $\sigma_Z = 0.16$ (first row) resulting in a rejection rate of 0.050. Each combination of bandwidths has the correct type 1 error rate of around $\alpha = 0.05$.

		$\sigma_X =$							
		0.72	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.69	0.046	0.046	0.049	0.042	0.049	0.050	0.048	0.047
	0.10	0.047	0.039	0.044	0.047	0.052	0.043	0.045	0.052
	0.20	0.050	0.038	0.045	0.049	0.049	0.055	0.051	0.049
	0.00	0.044	0.047	0.047	0.050	0.049	0.050	0.049	0.048
	1.00	0.052	0.046	0.043	0.050	0.047	0.050	0.052	0.050
	2.00	0.050	0.046	0.050	0.052	0.047	0.052	0.050	0.054
	5.00	0.042	0.046	0.044	0.046	0.045	0.048	0.047	0.051
	10.000	0.049	0.045	0.048	0.045	0.047	0.051	0.052	0.057

Table 13: Type 1 error rate of the kernel log-rank test with various combinations of bandwidths for $D.20$ of Table 4 when 75% of events is observed. The median heuristic yields $\sigma_X = 0.72$ (first column) and $\sigma_Z = 0.69$ (first row) resulting in a rejection rate of 0.050. Each combination of bandwidths has the correct type 1 error rate of around $\alpha = 0.05$ - none are above 0.57.

		$\sigma_X =$							
		2.94	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.49	0.047	*0.097*	*0.183*	0.047	0.044	0.046	0.046	0.047
	0.10	0.048	*0.120*	*0.206*	0.018	0.034	0.054	0.051	0.047
	0.20	0.048	*0.108*	*0.192*	0.028	0.038	0.045	0.053	0.046
	0.50	0.047	*0.102*	*0.172*	0.046	0.043	0.046	0.049	0.052
	1.00	0.045	*0.085*	*0.150*	0.049	0.046	0.051	0.052	0.049
	2.00	0.049	*0.078*	*0.131*	0.050	0.044	0.049	0.050	0.052
	5.00	0.052	*0.070*	*0.129*	0.053	0.047	0.055	0.053	0.053
	10.00	0.048	*0.079*	*0.120*	0.047	0.049	0.046	0.046	0.053

Table 14: Type 1 error rate of the kernel log-rank test with various combinations of bandwidths for $D.23$ of Table 4 when 75% of events is observed. The median heuristic yields $\sigma_X = 2.94$ (first column) and $\sigma_Z = 0.49$ (first row) resulting in a rejection rate of 0.053. We note that bandwidths σ_X which are small compared to the median heuristic of 2.94 lead to inflated type 1 error rate. Type 1 error rates above 0.57 are displayed in italics.

A.6.4 Power of the kernel (Gau,Gau) for different bandwidths

		$\sigma_X =$							
		0.72	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.69	0.86	0.44	0.65	0.80	0.86	0.90	0.89	0.89
	0.10	0.64	0.27	0.39	0.62	0.68	0.71	0.70	0.73
	0.20	0.77	0.37	0.55	0.74	0.81	0.84	0.83	0.82
	0.50	0.86	0.48	0.67	0.84	0.88	0.88	**0.91**	**0.91**
	1.00	0.76	0.37	0.50	0.70	0.81	0.83	0.82	0.82
	2.00	0.29	0.15	0.22	0.25	0.31	0.35	0.37	0.40
	5.00	0.15	0.10	0.10	0.13	0.13	0.17	0.16	0.18
	10.00	0.11	0.09	0.10	0.11	0.12	0.14	0.15	0.15

Table 15: Power of the kernel log-rank test with various combinations of bandwidths for $D.26$ of Table 7 when 75% of events is observed. The median heuristic yields $\sigma_X = 0.72$ (first column) and $\sigma_Z = 0.69$ (first row) resulting in a rejection rate of 0.86. The highest rejection rate 0.91 (in bold) is achieved when $(\sigma_Z, \sigma_X) = (0.5, 5), (0.5, 10)$. In D.19 we found that for this censoring distribution, each of the bandwidths led to a correct type 1 error rate.

		$\sigma_X =$							
		0.72	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.67	0.87	0.43	0.62	0.82	0.91	**0.95**	**0.95**	0.92
	0.10	0.68	0.30	0.41	0.64	0.74	0.76	0.79	0.78
	0.20	0.82	0.43	0.58	0.77	0.87	0.88	0.90	0.88
	0.50	0.90	0.47	0.69	0.85	0.92	0.94	**0.95**	0.94
	1.00	0.77	0.31	0.48	0.70	0.82	0.89	0.88	0.88
	2.00	0.29	0.11	0.14	0.24	0.34	0.39	0.36	0.37
	5.00	0.09	0.07	0.08	0.07	0.09	0.07	0.10	0.09
	10.00	0.06	0.06	0.07	0.06	0.06	0.08	0.07	0.07

Table 16: Power of the kernel log-rank test with various combinations of bandwidths for $D.27$ of Table 7 when 75% of events is observed. The median heuristic yields $\sigma_X = 0.72$ (first column) and $\sigma_Z = 0.67$ (first row) resulting in a rejection rate of 0.87. The highest rejection rate 0.94 (in bold) is achieved when $(\sigma_Z, \sigma_X) = (0.67, 2), (0.67, 5), (0.5, 5)$. In D.20 we found that for this censoring distribution, each of the bandwidths led to a correct type 1 error rate.

		$\sigma_X =$							
		2.96	0.10	0.20	0.50	1.00	2.00	5.00	10.00
$\sigma_Z =$	0.32	0.48	*0.07*	*0.12*	0.08	0.42	0.57	0.26	0.13
	0.10	0.34	*0.08*	*0.15*	0.03	0.28	0.41	0.20	0.10
	0.20	0.43	*0.08*	*0.11*	0.06	0.42	0.49	0.24	0.13
	0.50	0.55	*0.07*	*0.09*	0.06	0.50	0.64	0.31	0.14
	1.00	0.62	*0.07*	*0.10*	0.10	0.55	0.70	0.35	0.17
	2.00	0.63	*0.04*	*0.07*	0.08	0.58	0.73	0.37	0.19
	5.00	0.67	*0.05*	*0.07*	0.08	0.60	0.75	0.43	0.17
	10.00	0.67	*0.05*	*0.07*	0.09	0.60	**0.76**	0.41	0.19

Table 17: Power of the kernel log-rank test with various combinations of bandwidths for D.30 7 when 75% of events is observed. The median heuristic yields $\sigma_X = 2.96$ (first column) and $\sigma_Z = 0.32$ (first row) corresponding to a rejection rate of 0.48. The highest rejection rate 0.76 is achieved when $(\sigma_Z, \sigma_X) = (10, 2)$. Rejection rates of bandwidths that led to an increased type 1 error rate in D.23, which featured the same censoring distribution, are displayed in italics.

B Preliminary results

In this section, and in order for this paper to be self-contained, we review some preliminary results that will be used on our proofs.

B.1 Counting processes

We state some results from the counting process theory that are frequently used in our paper. Recall that $Y(t) = \sum_{i=1}^n Y_i(t)$ denotes the pooled risk function, where $Y_i(t) = \mathbb{1}_{\{T_i \geq t\}}$, $\tau = \sup\{t : S_T(t) > 0\}$, $\tau_n = \max\{T_1, \dots, T_n\}$, and that Z and C are continuous random variables.

Proposition B.1. *The following holds a.s.*

1. $\lim_{n\to\infty} \sup_{t\leq\tau} |Y(t)/n - S_T(t)| = 0$
2. $\sup_{t\leq t^\star} \left|\frac{n}{Y(t)} - \frac{1}{S_T(t)}\right| \to 0$. *for any* $t^\star \in (0, \tau)$

The proof of Part 1. is due to Glivenko–Cantelli's Theorem, and Part 2. is a consequence of Part 1. Also, notice that under the null hypothesis $S_T(t) = S_Z(t)\int_{\mathbb{R}^d} S_{C|X=x}(t)dF_X(x)$ since we assume non-informative censoring, i.e. $Z \perp C|X$.

Proposition B.2. *Let* $\beta \in (0, 1)$*, then*

1. $\mathbb{P}\left(Y(t)/n \leq \beta^{-1}S_T(t), \quad \forall t \leq \tau_n\right) \geq 1 - \beta$,
2. $\mathbb{P}\left(Y(t)/n \geq \beta S_T(t), \quad \forall t \leq \tau_n\right) \geq 1 - e(1/\beta)e^{-1/\beta}$.

The proof of Part 1. follows from [12, Theorem 3.2.1.] and Part 2. is due [13].

B.2 Dominated convergence in probability

Lemma B.3. *Let* $(\mathcal{X}, \mathcal{B}, \mu)$ *be a measurable space and* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Consider a sequence of non-negative stochastic processes* $R_n : \Omega \times \mathcal{X} \to \mathbb{R}$*. Suppose that*

i) For each $\alpha \in (0, 1)$*, there exists an event* A_α *with* $\mathbb{P}(A_\alpha) \geq 1 - \alpha$*, such that* $R_n(\omega, x) \to 0$ *for all* $(\omega, x) \in A_\alpha \times \mathcal{X}$*.*

ii) For each $\beta \in (0, 1)$*, there exists a non-negative function* $R_\beta \in L_1(\mathcal{X}, \mathcal{B}, \mu)$ *and* N_0 *large enough such that for each* $n \geq N_0$ *there exists an event* $B_{n,\beta}$ *with* $\mathbb{P}(B_{n,\beta}) \geq 1 - \beta$ *and*

$$R_n(\omega, x) \leq R_\beta(x)$$

for $(\omega, x) \in B_{n,\beta} \times \mathcal{X}$*.*

Then $I_n(\omega) = \int_{\mathcal{X}} R_n(\omega, x)\mu(dx) \to 0$ *in* $\mathbb{P}$*-probability.*

The proof follows by noticing that in a set of probability tending to one, we can apply dominated convergence.

B.3 Lenglart-Rebolledo Inequality

Let $X(t)$ and $X'(t)$ be right-continuous adapted processes. We say that X is majorised by X' if for all bounded stopping times T it holds that $\mathbb{E}(|X(T)|) \leq \mathbb{E}(X'(T))$.

Lemma B.4 (Theorem 3.4.1. of [10])**.** *Let* X *be a right-continuous adapted process, and* X' *a non-decreasing predictable process with* $X'(0) = 0$ *such that* X *is majorised by* X'*. For any stopping time* T*, and any* $\varepsilon, \delta > 0$*,*

$$\mathbb{P}\left(\sup_{t\leq T} |X(t)| > \varepsilon\right) \leq \frac{\delta}{\varepsilon} + \mathbb{P}(X'(T) \geq \delta)$$

In our setting $X(t)$ is a sub-martingale $W_n(t)$ (based on n data points) and $X'(t)$ is the corresponding compensator $A_n(t)$. We consider the stopping time $\tau_n = \max\{T_1, \dots, T_n\}$, and apply the previous Lemma to prove

$$A_n(\tau_n) = o_p(1) \Rightarrow \sup_{t\leq\tau_n} W_n(t) = o_p(1). \tag{17}$$

B.4 Double martingale integrals

Next we state results regarding double integrals with respect to a $\mathcal{F}_t$- counting process martingale $M(t)$, introduced in [8]. Consider $W(t) = \int_{C_t} h(x,y)dM(x)dM(y)$, where $C_t = \{(x,y) : 0 < x < y \leq t\}$. The following results state conditions under which $W(t)$ defines a proper martingale with respect the filtration $(\mathcal{F}_t)_{t\geq 0}$.

Definition B.5. We define the predictable σ- algebra $\mathcal{P}$ as the σ-algebra generated by the sets of the form

$$\sigma\{\{(a_1,b_1] \times (a_2,b_2] \times X : 0 \leq a_1 \leq b_1 < a_2 \leq b_2, X \in \mathcal{F}_{a_2}\} \cup \{(0,0) \times X : X \in \mathcal{F}_0\}\}.$$

Let $C = \{(x,y) : 0 < x < y < \infty\}$. A function $h : C \times \Omega \to \mathbb{R}$ is called *"elementary predictable"* if it can be written as a finite sum of indicator functions of sets belonging to the predictable σ-algebra $\mathcal{P}$. On the other hand, if a function h is $\mathcal{P}$-measurable, then it is the limit of elementary predictable functions.

Proposition B.6. *Any deterministic function, and the functions $C \times \Omega \to \mathbb{R}$: $Y(x)Y(y)$ and $\mathbb{1}_{\{x<y\leq\tau_n\}}$ are $\mathcal{P}$-measurable.*

Remark. Notice that any deterministic function of the covariates $X_1, \ldots, X_n$, individual risk functions $Y_i(t)$ and indicator function $\mathbb{1}_{\{t\leq\tau_n\}}$ are $\mathcal{P}$-measurable as in Definition B.5.

Theorem B.7. *Let $h(x,y)$ be $\mathcal{P}$-measurable function (see Definition B.5) and define $C_t = \{(x,y) : 0 < x < y \leq t\}$ with $t \geq 0$. Let $M(t)$ be a $\mathcal{F}_t$-martingales and suppose that for all $t \geq 0$, it holds*

$$\mathbb{E}\left(\int_{C_t} |h||dM(x)dM'(y)|\right) < \infty. \tag{18}$$

Then, the process

$$W(t) = \int_{C_t} h(x,y)dM(x)dM'(y)$$

is a martingale on $t \geq 0$ with respect the filtration $\mathcal{F}_t$.

B.5 Forward and Backward operators

Consider the forward and backward operators introduced in [7], $A : \mathcal{L}_2(Z \times C \times X) \to \mathcal{L}_2(Z \times C \times X)$, and $B : \mathcal{L}_2(Z \times C \times X) \to \mathcal{L}_2(Z \times C \times X)$, respectively, defined as

$$(Af)(z,c,x) = f(z,c,x) - \frac{1}{S_Z(z)}\int_z^\infty f(s,c,x)dF_Z(s)$$

and

$$(Bf)(z,c,x) = f(z,c,x) - \int_0^z f(s,c,x)d\Lambda_Z(s).$$

It was proved in [7] that the previous operators satisfy the following properties. For any $f, g \in \mathcal{L}_2(Z \times C \times X)$:

1. $\mathbb{E}((Af)g|C_1, X_1) = \mathbb{E}(f(Bg)|C_1, X_1)$
2. $ABf = f$
3. $BAf = f - \mathbb{E}(f|C_1, X_1)$
4. $\mathbb{E}(Bf|C_1, X_1) = 0$
5. For any $f(z,c,x) = \mathbb{1}_{\{z\leq c\}}g(z,c,x)$,

$$(Bf)(z,c,x) = \int_{\mathbb{R}_+} g(s,c,x)dm_{z,c}(s),$$

where $dm_{z,c}(s) = \mathbb{1}_{z\leq c}\delta_z(s) - \mathbb{1}_{\{\min\{z,c\}\geq s\}}d\Lambda_Z(s)$.

C Auxiliary results

In this section we introduce some auxiliary results that we will be used in the proofs of our main results. The proofs of these results are given in Section E.

- Proposition C.1 is a technical result that ensures τ is consistent with the definition of τ_x.
- Proposition C.2 is used in the proof of Theorem 3.1.
- Lemma C.3 is used in the proof of Lemma C.4.
- Lemma C.4 is used in the proof of the main result, Lemma 4.3.
- Theorem C.5 is used in the proof of Theorem 4.5
- Lemma C.6 is used in the proof of Lemma 4.6.

Proposition C.1. *Let τ' be the essential supremum defined as $\tau' = ess\ \sup \tau_x = \inf\{t : F_X(\{x : \tau_x > t\}) = 0\}$, then $\tau = \tau'$.*

Proposition C.2. *Let $\omega : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function. Then*

$$\int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)d\nu_0^n(s,x) = \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n \Delta_j\omega(T_j,X_i)\frac{Y_i(T_j)}{S_T(T_j)} + o_p(1).$$

Lemma C.3. *Consider Assumption 4.2 holds (L and K denote kernels on the times and covariates, respectively), and let $\kappa_{i,j} : \Omega \times (0,\tau)^2 \to \mathbb{R}$ be a process given by*

$$\begin{aligned}\kappa_{i,j}(t,s) = &\frac{\mathbb{E}'(K(X_i',X_j')Y_i'(t)Y_j'(s))}{n^2S_T(t)S_T(s)} - \frac{\mathbb{E}'(K(X_i',X_j)Y_i'(t))Y_j(s)}{nS_T(t)Y(s)}\\ &- \frac{\mathbb{E}'(K(X_i,X_j')Y_j'(s))Y_i(t)}{nS_T(s)Y(t)} + \frac{K(X_i,X_j)Y_i(t)Y_j(s)}{Y(t)Y(s)},\end{aligned}$$

where $((T_i',\Delta_i',X_i'))_{i=1}^n$ is an i.i.d. independent sample of the data $\mathcal{D} = ((T_i,\Delta_i,X_i))_{i=1}^n$, $Y_i'(t) = \mathbb{1}\{T_i' \geq t\}$, and $\mathbb{E}'(\cdot) = \mathbb{E}(\cdot|\mathcal{D})$.

Then,

$$\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t) = o_p(1),$$

point-wise for any $t < \tau$.

Lemma C.4. *Consider Assumption 4.2 holds, and let $\kappa_{i,j}$ be the process defined in Lemma C.3. Then, the following results hold true:*

i) $\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t) = o_p(1)$

ii) $\left|\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t)\right| \leq C$*, for some constant $C > 0$.*

Theorem C.5. *Let $Q : (\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}^d)^2 \to \mathbb{R}$, $T^Q : \mathcal{L}_2(Z\times C\times X) \to \mathcal{L}_2(Z\times C\times X)$ and $T^J : \mathcal{L}_2(T\times\Delta\times X) \to \mathcal{L}_2(T\times\Delta\times X)$ be given by*

$$Q((z_1,c_1,x_1),(z_2,c_2,x_2)) = \mathbb{1}_{\{z_1\leq c_1\}}\mathbb{1}_{\{z_2\leq c_2\}}\bar{\mathfrak{K}}((z_1,x_1),(z_2,x_2)),$$

$$(T^Qf)(Z_1,C_1,X_1) = \mathbb{E}_2\left(Q((Z_1,C_1,X_1),(Z_2,C_2,X_2))f(Z_2,C_2,X_2)\right),$$

and

$$(T^Jf)(T_1,\Delta_1,X_1) = \mathbb{E}_2\left(J((T_1,\Delta_1,X_1),(T_2,\Delta_2,X_2))f(T_2,\Delta_2,X_2)\right), \tag{19}$$

respectively, where J is the function defined in Equation (12) in the main document, $\mathbb{E}_2(\cdot) = \mathbb{E}(\cdot|(T_1,\Delta_1,X_1))$, and (Z_1,C_1,X_1) and (Z_2,C_2,X_2) are independent copies of our random variables in the augmented space.

Then, $T^J = BT^QA$, and T^J and T^Q have the same set of non-zero eigenvalues, including multiplicities.

Lemma C.6.

$$\frac{1}{n}\int_0^\tau\left|1 - \frac{Y(t)/n}{S_T(t)}\right|^k dN(t) = o_p(1), \tag{20}$$

for $k = 1$ and $k = 2$.

D Main proofs

In this section we prove the main results of our paper.

D.1 Proofs of Section 3

D.1.1 Proof of Theorem 3.1

Proof: The result follows from proving

$$\int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)d\nu_k^n(s,x) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)d\nu_k(t,x)$$

for $k \in \{0,1\}$.

The result for $k=1$ follows from a direct application of the law of large numbers. We continue with the result for $k=0$.

By Proposition C.2, it holds

$$\int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)d\nu_0^n(s,x) = \frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n \Delta_j\omega(T_j,X_i)\frac{Y_i(T_j)}{S_T(T_j)} + o_p(1).$$

The sum in the left-hand-side of the previous equation can be decomposed into two sums, one over indices such that $\{i \neq j\}$ and the other over indices such that $\{i=j\}$. Notice that the sum over $\{i\neq j\}$ satisfies

$$\frac{1}{n^2}\sum_{i\neq j}^n \Delta_j\omega(T_j,X_i)\frac{Y_i(T_j)}{S_T(T_j)} \xrightarrow{\mathbb{P}} \mathbb{E}\left(\Delta_2\omega(T_2,X_1)\frac{Y_1(T_2)}{S_T(T_2)}\right) = \int_{\mathbb{R}^d}\int_0^\tau \omega(s,x)d\nu_0(s,x)$$

by the law of large numbers for U-statistics. Thus, to conclude the proof, we just need to prove that the sum over indices $\{i=j\}$ converges to zero in probability. To see this, notice that $U_i = S_T(T_i)$ are i.i.d. uniform random variables, and that

$$\frac{1}{n^2}\sum_{i=1}^n \Delta_i\omega(T_i,X_i)\frac{1}{S_T(T_i)} \leq \frac{C}{n^2}\sum_{i=1}^n\frac{1}{S_T(T_i)} \leq \frac{C}{n^2}\sum_{i=1}^n \frac{1}{U_i} = o_p(1),$$

where the first inequality follows from the assumption ω is bounded by some constant C, and the last equality follows from proving $n^{-2}\sum_{i=1}^n U_i^{-1} = o_p(1)$, which we proceed to prove.

Define the event $\mathcal{B}_n = \cup_{i=1}^n\left\{1/U_i \geq n^{3/2}\right\}$, and notice that $\mathbb{P}(\mathcal{B}_n) \leq n\mathbb{P}(U_i \leq 1/n^{3/2}) = n^{-1/2}$. Then, for any $\epsilon > 0$, it holds

$$\begin{aligned}
\mathbb{P}\left(\frac{C}{n^2}\sum_{i=1}^n\frac{1}{U_i} > \epsilon\right) &= \mathbb{P}\left(\left\{\frac{C}{n^2}\sum_{i=1}^n\frac{1}{U_i} > \epsilon\right\}\cap\mathcal{B}_n\right) + \mathbb{P}\left(\left\{\frac{C}{n^2}\sum_{i=1}^n\frac{1}{U_i} > \epsilon\right\}\cap\mathcal{B}_n^c\right)\\
&\leq \mathbb{P}(\mathcal{B}_n) + \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n\frac{\mathbb{1}_{\{1/U_i<n^{3/2}\}}}{U_i} > \frac{n\epsilon}{C}\right)\\
&\leq \frac{1}{n^{1/2}} + \frac{C}{n\epsilon}\mathbb{E}\left(\frac{\mathbb{1}_{\{1/U_i<n^{3/2}\}}}{U_i}\right)\\
&\leq \frac{1}{n^{1/2}} + \frac{C}{n\epsilon}\int_{n^{-3/2}}^1\frac{1}{u}du\\
&\leq \frac{1}{n^{1/2}} + \frac{3C}{2\epsilon}\frac{\log(n)}{n} \to 0,
\end{aligned}$$

for any $\epsilon > 0$ as n grows to infinity.

■

D.1.2 Proof of Proposition 3.3

Proof: ($\Longleftarrow$) Assume $Z \perp X$. Notice that $\tau = \tau_x$ for almost all $x \in \mathbb{R}^d$ as if this does not hold, then $Z \not\perp X$. Consider arbitrary $t \leq \tau$ and measurable $A \subseteq \mathbb{R}^d$, then it holds

$$\nu_1((0,t] \times A) = \int_0^t \int_{x \in A} S_{C|X=x}(s) dF_Z(s) dF_X(x),$$

and

$$\begin{aligned}\nu_0((0,t] \times A) &= \int_0^t \int_{x \in \mathbb{R}^d} S_{C|X=x}(s) \frac{\int_{x' \in A} S_Z(s) S_{C|X=x'}(s) dF_X(x')}{\int_{x'' \in \mathbb{R}^d} S_Z(s) S_{C|X=x''}(s) dF_X(x'')} dF_Z(s) dF_X(x) \\ (Z \perp X) &= \int_0^t \int_{x \in A} S_{C|X=x}(s) dF_Z(s) dF_X(x).\end{aligned}$$

Because ν_1 and ν_0 agree on generating sets, they agree on any measurable set of $\mathbb{R}_+ \times \mathbb{R}^d$.

($\Longrightarrow$) First, we prove that $\nu_1 = \nu_0$ implies that $\tau = \tau_x$ for almost all $x \in \mathbb{R}^d$. Observe that $S_{C|X=x}(s) = 0 \Rightarrow S_{Z|X=x}(s) = 0$ implies $\tau_x = \sup\{t : S_{T|X=x}(t) > 0\} = \sup\{t : S_{Z|X=x}(t) > 0\}$, and by Proposition C.1, $\tau_x \leq \tau$ for almost all $x \in \mathbb{R}^d$.

We proceed by contradiction. Assume there exists $\tau' < \tau$ such that the event $E = \{x : \tau_x \leq \tau'\}$ satisfies $\mathbb{P}(X \in E) \in (0,1)$, otherwise $\tau_x = \tau$ for almost all $x \in \mathbb{R}^d$. Let $\omega(s,x) = \mathbb{1}_{\{s<\tau_x\}} S_{T|X=x}(s)^{-1}$, and notice that $\omega(s,x) < \infty$ for all $(s,x) \in (0,\tau) \times \mathbb{R}^d$. By the hypothesis $\nu_0 \equiv \nu_1$,

$$\int_A \int_0^t \omega(s,x) S_{C|X=x}(s) dF_{ZX}(s,x) = \int_A \int_0^t \omega(s,x) S_{T|X=x}(s) d\alpha(s) dF_X(x) \tag{21}$$

for any $t \leq \tau$, measurable set $A \subseteq \mathbb{R}^d$, and measurable function $\omega(s,x)$.

The left-hand-side of equation (21), when evaluated at $\omega(s,x) = \mathbb{1}_{\{s<\tau_x\}} S_{T|X=x}(s)^{-1}$, $A = E$ and $t = \tau'$, then satisfies

$$\begin{aligned}\int_E \int_0^{\tau'} \omega(s,x) S_{C|X=x}(s) dF_{ZX}(s,x) &= \int_E \int_0^{\tau'} \mathbb{1}_{\{s<\tau_x\}} d\Lambda_{Z|X=x}(s) dF_X(x) \\ (\tau_x < \tau' \text{ in } E) &= \int_E \Lambda_{Z|X=x}(\tau_x) d\mathcal{F}_X(x) \\ &= \infty,\end{aligned}$$

since $\tau_x = \sup\{t : S_{Z|X=x}(t) > 0\}$, $\Lambda_{Z|X=x}(t) = -\log S_{Z|X=x}(t)$ and $F_X(E) > 0$.

On the other hand, the right-hand-side of equation (21), when evaluated at $\omega(s,x) = \mathbb{1}_{\{s<\tau_x\}} S_{T|X=x}(s)^{-1}$, $A = E$ and $t = \tau'$, satisfies

$$\begin{aligned}\int_E \int_0^{\tau'} \omega(s,x) S_{T|X=x}(s) d\alpha(s) dF_X(x) &= \int_E \int_0^{\tau'} \mathbb{1}_{\{s\leq\tau_x\}} d\alpha(s) dF_X(x) \\ &\leq F_X(E)\alpha(\tau') < \infty,\end{aligned}$$

since $F_X(E) \in (0,1)$ and

$$\alpha(\tau') = \int_0^{\tau'} \frac{\int_{x \in \mathbb{R}^d} S_{C|X=x}(s) dF_{ZX}(s,x)}{S_T(s)} \leq \frac{1}{S_T(\tau')} < \infty$$

due to the fact that $S_T(t) > 0$ for all $t < \tau$ and $\tau' < \tau$. This finally, leads to a contradiction in equation (21), hence $\tau_x = \tau$ for almost all $x \in \mathbb{R}^d$.

We continue assuming that $\tau = \tau_x$ for almost all $x \in \mathbb{R}^d$ and prove that $\nu_0 = \nu_1$ implies $Z \perp X$. Let $\omega(s,x) = S_{T|X=x}(s)^{-1}$, $A \subseteq \mathbb{R}^d$ be a measurable set and $t < \tau$. Notice that ω is well-defined F_X-almost surely for $t < \tau$ as $F_X(\{\tau_X \neq \tau\}) = 0$. By the hypothesis $\nu_0 \equiv \nu_1$,

$$\int_0^t \int_A \frac{S_{C|X=x}(s)}{S_{T|X=x}(s)} dF_{ZX}(s,x) = \int_0^t \int_A d\alpha(s) dF_X(x) \tag{22}$$

By replacing the value of ω and choosing $A = \mathbb{R}^d$, we obtain

$$\int_{\mathbb{R}^d} \int_0^t \frac{dF_{Z|X=x}(s)}{S_{Z|X=x}(s)} dF_X(x) = \int_{\mathbb{R}^d} \Lambda_{Z|X=x}(t) dF_X(x) = \Lambda_Z(t) = \int_0^t d\alpha(t).$$

which implies $d\alpha(s) = d\Lambda_Z(s)$. Substituting the previous equality in equation (22), we obtain for arbitrary measurable A,

$$\int_0^t \int_A \frac{S_{C|X=x}(s)}{S_{T|X=x}(s)} dF_{ZX}(s,x) = \int_A \Lambda_{Z|X=x}(t) dF_X(x) = \int_A \Lambda_Z(t) dF_X(x)$$

from which we conclude that $d\Lambda_{Z|X}(s) = d\Lambda_Z(s)$ F_X-almost surely for all $s < \tau$. Notice that $S_{C|X=x}(s) = 0 \Rightarrow S_{Z|X=x}(s) = 0$ implies that $\tau = \sup\{t : S_Z(t) > 0\}$ from which we deduce $Z \perp X$. ■

D.1.3 Proof of Theorem 3.4

Proof: Let $\omega^\star = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^{-1}(\phi_0^n - \phi_1^n)$, clearly $\|\omega^\star\|_{\mathcal{H}}^2 = 1$ and

$$\langle \omega^\star, \phi_0^n - \phi_1^n \rangle_{\mathcal{H}} = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}} \leq \Psi_n.$$

On the other hand

$$\Psi_n = \sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \leq 1} \langle \omega, \phi_0^n - \phi_1^n \rangle_{\mathcal{H}} \leq \sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \leq 1} \|\omega\|_{\mathcal{H}} \|\phi_0^n - \phi_1^n\|_{\mathcal{H}} = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}},$$

which deduces the desired result $\Psi_n = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}$.

Observe

$$\begin{aligned}
\|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 &= \left\| \int_{\mathbb{R}^d} \int_{\mathbb{R}_+} L(t,\cdot)K(x,\cdot)(d\nu_1^n(t,x) - d\nu_0^n(t,x)) \right\|_{\mathcal{H}}^2 \\
&= \left\| \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}_+} L(t,\cdot) \left(K(X_i,\cdot) - \sum_{j=1}^n K(X_j,\cdot) \frac{Y_j(t)}{Y(t)} \right) dN_i(t) \right\|_{\mathcal{H}}^2 \\
&= \left\| \frac{1}{n} \sum_{i=1}^n \Delta_i L(T_i,\cdot) \left(K(X_i,\cdot) - \sum_{j=1}^n K(X_j,\cdot) \frac{Y_j(T_i)}{Y(T_i)} \right) \right\|_{\mathcal{H}}^2.
\end{aligned}$$

Define $\boldsymbol{\Phi}^{\boldsymbol{X}} = (K(X_1,\cdot),\dots,K(X_n,\cdot))^\intercal$ and $\boldsymbol{\Phi}^{\boldsymbol{T}} = (\Delta_1 L(T_1,\cdot),\dots,\Delta_n L(T_n,\cdot))^\intercal$, then

$$\begin{aligned}
\|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 &= \left\| \frac{1}{n} \sum_{i=1}^n \Phi_i^T ((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{\Phi}^{\boldsymbol{X}})_i \right\|_{\mathcal{H}}^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n \left\langle \Phi_i^T ((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{\Phi}^{\boldsymbol{X}})_i, \Phi_{i'}^T ((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{\Phi}^{\boldsymbol{X}})_{i'} \right\rangle_{\mathcal{H}} \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n (\boldsymbol{L}^{\boldsymbol{\Delta}})_{i,i'} ((\boldsymbol{I} - \boldsymbol{A})\boldsymbol{K}(\boldsymbol{I} - \boldsymbol{A})^\intercal)_{i,i'} \\
&= \frac{1}{n^2} \operatorname{trace}(\boldsymbol{L}^{\boldsymbol{\Delta}}(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{K}(\boldsymbol{I} - \boldsymbol{A})^\intercal).
\end{aligned}$$

■

D.2 Proofs of Section 4

D.2.1 Proof of Proposition 4.1

Proof: Observe that

$$\begin{aligned}
(\phi_0^n - \phi_1^n)(\cdot) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}^d} \mathfrak{K}((t,x),\cdot)(d\nu_1^n(t,x) - d\nu_0^n(t,x)) \\
&= \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t,X_i),\cdot) - \sum_{j=1}^n \mathfrak{K}((t,X_j),\cdot) \frac{Y_j(t)}{Y(t)} \right) dN_i(t)
\end{aligned}$$

follows from the definition of the embeddings, ϕ_0^n and ϕ_1^n, and of the empirical measures, ν_1^n and ν_0^n.

Assume that the process N_i can be replaced by the martingale M_i in the previous equation, that is,

$$(\phi_0^n - \phi_1^n)(\cdot) = \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t, X_i), \cdot) - \sum_{j=1}^n \mathfrak{K}((t, X_j), \cdot) \frac{Y_j(t)}{Y(t)} \right) dM_i(t). \tag{23}$$

Then, using Theorem 3.4 and the previous result, it holds

$$\Psi_n^2 = \|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 = \left\| \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t, X_i), \cdot) - \sum_{j=1}^n \mathfrak{K}((t, X_j), \cdot) \frac{Y_j(t)}{Y(t)} \right) dM_i(t) \right\|_{\mathcal{H}}^2,$$

where the main result is deduced by computing the previous term using the reproducing kernel property.

To end the proof, we just need to prove Equation (23). Since $dM_i(t) = dN_i(t) - Y_i(t)d\Lambda_Z(t)$, the result follows from proving

$$\sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t, X_i), \cdot) - \sum_{j=1}^n \mathfrak{K}((t, X_j), \cdot) \frac{Y_j(t)}{Y(t)} \right) Y_i(t) d\Lambda_Z(t) = 0.$$

Observe

$$\begin{aligned}
&\sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t, X_i), \cdot) - \sum_{j=1}^n \mathfrak{K}((t, X_j), \cdot) \frac{Y_j(t)}{Y(t)} \right) Y_i(t) d\Lambda_Z(t) \\
&= \sum_{i=1}^n \int_{\mathbb{R}_+} \mathfrak{K}((t, X_i), \cdot) Y_i(t) d\Lambda_Z(t) - \sum_{i=1}^n \int_{\mathbb{R}_+} \sum_{j=1}^n \mathfrak{K}((t, X_j), \cdot) \frac{Y_j(t)}{Y(t)} Y_i(t) d\Lambda_Z(t) \\
&= \sum_{i=1}^n \int_{\mathbb{R}_+} \mathfrak{K}((t, X_i), \cdot) Y_i(t) d\Lambda_Z(t) - \sum_{j=1}^n \int_{\mathbb{R}_+} \mathfrak{K}((t, X_j), \cdot) Y_j(t) d\Lambda_Z(t) \\
&= 0
\end{aligned}$$

which completes the result.

■

D.2.2 Proof of Lemma 4.3

Proof: Recall the definition of $\mathrm{LR}_n(\omega)$ from Equation (3) in the main document. By following the same steps of the proof Proposition 4.1, it is easy to prove

$$\mathrm{LR}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}_+} (\omega(t, X_i) - \bar{\omega}_n(t)) dM_i(t),$$

and thus

$$\Psi_n = \sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \le 1} \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\omega(t, X_i) - \sum_{j=1}^n \omega(t, X_j) \frac{Y_j(t)}{Y(t)} \right) dM_i(t).$$

Let $\{(T_i', \Delta_i', X_i')\}_{i=1}^n$ be an independent copy of the data $\mathcal{D} = \{(T_i, \Delta_i, X_i)\}_{i=1}^n$, let $Y_i'(t) = \mathbb{1}_{\{T_i' \ge t\}}$ and let $\mathbb{E}'(\cdot) = \mathbb{E}(\cdot|\mathcal{D})$. The desired result is then obtained by showing that the term $\sum_{j=1}^n \omega(t, X_j) \frac{Y_j(t)}{Y(t)}$ can be replaced in the previous equation by its population limit, given by $\mathbb{E}(\omega(t, X_1'), Y_1'(t))/S_T(t)$, up to an error of order $o_p(n^{-1/2})$ (notice this is not obvious because of the supremum).

Indeed, let's make this replacement, and define

$$\begin{aligned}
\bar{\Psi}_n &= \sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \le 1} \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\omega(t, X_i) - \frac{\mathbb{E}(\omega(t, X_1')Y_1'(t))}{S_T(t)} \right) dM_i(t) \\
&= \sup_{\omega \in \mathcal{H}: \|\omega\|_{\mathcal{H}}^2 \le 1} \left\langle \omega, \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t, X_i), \cdot) - \frac{\mathbb{E}(\mathfrak{K}((t, X_1'), \cdot)Y_1'(t))}{S_T(t)} \right) dM_i(t) \right\rangle.
\end{aligned}$$

Since the supremum is taken over the unit ball of a reproducing kernel Hilbert space, it is straightforward that

$$\begin{aligned}\bar{\Psi}_n^2 &= \left\| \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}_+} \left(\mathfrak{K}((t,X_i),\cdot) - \frac{\mathbb{E}(\mathfrak{K}((t,X_1'),\cdot)Y_1'(t))}{S_T(t)} \right) dM_i(t) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \bar{\mathfrak{K}}((t,X_i),(t',X_l))dM_i(t)dM_l(t'),\end{aligned}$$

where $\bar{\mathfrak{K}}$ is the population limit of $\bar{\mathfrak{K}}_n$ (under the null) given in Equation (11) in the main document. Thus, the desired result follows from proving $\Psi_n^2 = \bar{\Psi}_n^2 + o_p(n^{-1})$.

Define the error term

$$\Upsilon_n = \sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}^2\leq 1} \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}_+} \sum_{j=1}^n \left(\frac{\mathbb{E}(\omega(t,X_j'),Y_j'(t))}{nS_T(t)} - \omega(t,X_j)\frac{Y_j(t)}{Y(t)} \right) dM_i(t),$$

and notice that, by the triangular inequality, it holds $\Psi_n \leq \bar{\Psi}_n + \Upsilon_n$ and $\Psi_n \geq \bar{\Psi}_n - \Upsilon_n$. If we assume that $\Upsilon_n = o_p(n^{-1/2})$ holds, the result follows by taking square of

$$n\Psi_n^2 = (n^{1/2}\bar{\Psi}_n + o_p(1))^2 = n\bar{\Psi}_n^2 + o_p(1),$$

notice that $n^{1/2}\bar{\Psi}_n o_p(1) = o_p(1)$ since $n^{1/2}\bar{\Psi}_n$ converges in distribution (which is proved in Theorem 4.4) to some random variable.

The rest of this proof is concerned with proving $\Upsilon_n = o_p(n^{-1/2})$ holds under Assumption 4.2. Again, observe that since the supremum in Υ_n is taken over the unit ball of a reproducing kernel Hilbert space, Υ_n^2 satisfies

$$\Upsilon_n^2 = \left\| \frac{1}{n}\int_{\mathbb{R}_+} \sum_{j=1}^n \left(\frac{\mathbb{E}(\mathfrak{K}((t,X_j'),\cdot),Y_j'(t))}{nS_T(t)} - \mathfrak{K}((t,X_j),\cdot)\frac{Y_j(t)}{Y(t)} \right) dM_j(t) \right\|_{\mathcal{H}}^2. \tag{24}$$

Under Assumption 4.2, $\mathfrak{K}((t,x),(t',x')) = L(t,t')K(x,x')$, a simple computation shows

$$\Upsilon_n^2 = \frac{1}{n^2}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \sum_{i=1}^n\sum_{j=1}^n L(t,t')\kappa_{i,j}(t,t')dM(t)dM(t'), \tag{25}$$

where $\kappa_{i,j}$ is the process defined in Lemma C.3. By Markov's inequality, for any $\delta > 0$, it holds

$$\mathbb{P}(n\Upsilon_n^2 > \delta) \leq \frac{E(n\Upsilon_n^2)}{\delta}. \tag{26}$$

Proving $\Upsilon_n = o_p(n^{-1/2})$ is thus equivalent to proving that $\mathbb{E}(n\Upsilon_n^2)$ converges to zero. To this end, decompose Υ_n^2 as

$$\begin{aligned}\Upsilon_n^2 = \frac{1}{n^2}\int_0^{\tau_n}\int_0^{\tau_n} \sum_{i=1}^n\sum_{j=1}^n \mathbb{1}_{\{t\neq t'\}} L(t,t')\kappa_{i,j}(t,t')dM(t)dM(t') \\ + \frac{1}{n^2}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t),\end{aligned} \tag{27}$$

where the first integral is considered over the off-diagonal $\{t \neq t'\}$ and the second integral is considered over the diagonal $\{t = t'\}$. Notice that for the integral over the diagonal $(dM(t))^2 = dN(t)$ follows from the fact we are considering continuous survival times T_i.

We proceed to prove that the expectation over the off-diagonal term converges to zero. Define the process $(Z(s))_{s\geq 0}$ by

$$Z(s) = \frac{2}{n^2}\int_0^s\int_{(0,t)}\sum_{i=1}^n\sum_{j=1}^n L(t,t')\kappa_{i,j}(t,t')dM(t')dM(t),$$

and notice that, by the symmetry of the functions $\kappa_{i,j}(t,t')$ and $L(t,t')$, the off-diagonal term in Υ_n^2 equals $Z(\tau_n)$, i.e,

$$\frac{1}{n^2}\int_0^{\tau_n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n \mathbb{1}_{\{t\neq t'\}} L(t,t')\kappa_{i,j}(t,t')dM(t')dM(t) = Z(\tau_n).$$

On the other hand, by Theorem B.7, the process $Z(t)$ is a zero-mean square-integrable $\mathcal{F}_t$- martingale since the functions $L(t,t')$ and $\kappa_{i,j}(t,t')$ (for any $i,j \in \{1,\dots,n\}$) are both predictable functions in the sense of definition B.5. By an application of the optional stopping theorem, it follows that $E(Z(\tau_n)) = 0$, and thus

$$\mathbb{E}(n\Upsilon_n^2) = \mathbb{E}\left(\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t)\right). \tag{28}$$

We conclude our proof by showing that the previous expectation is zero. To see this, we use the following results, proved in Lemma C.4 under Assumption 4.2:

i) $\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t) = o_p(1)$

ii) $\left|\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t)\right| \leq C$, for some $C > 0$,

From i), for any given $\epsilon > 0$, there exists $N_0 \geq 0$ large enough such that the event

$$\mathcal{B}_\epsilon^n = \left\{\left|\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t)\right| \leq \epsilon/2\right\} \tag{29}$$

satisfies $\mathbb{P}(\mathcal{B}_\epsilon^n) \geq 1 - \epsilon(2C)^{-1}$ for all $n \geq N_0$, where $C > 0$ is the constant defined in ii). Combining i) and ii) and equation (28), we obtain

$$\mathbb{E}(n\Upsilon_n^2) = \mathbb{E}(n\Upsilon_n^2\mathbb{1}_{\{\mathcal{B}_\epsilon^n\}}) + \mathbb{E}(n\Upsilon_n^2\mathbb{1}_{\{\mathcal{B}_\epsilon^n\}^c}) \leq \epsilon/2 + C\mathbb{P}(\{\mathcal{B}_\epsilon^n\}^c) = \epsilon, \tag{30}$$

which implies $\mathbb{E}(n\Upsilon_n^2) \to 0$ as n grows to infinity. The previous result combined with Equation (26) deduced the desired result, $n\Upsilon_n^2 = o_p(1)$. ■

D.2.3 Proof of Theorem 4.4

Proof: By Lemma 4.3,

$$\Psi_n^2 = \frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n J((T_i,\Delta_i,X_i),(T_l,\Delta_l,X_l)) + o_p(n^{-1})$$

under the null hypothesis. The non-negligible part in the right-hand side of the previous equation is a V-statistic of order 2. The degeneracy property can be easily verified since $\mathbb{E}(J(T_i,\Delta_i,X_i),(t,r,x)) = 0$ for any $(t,r,x) \in \mathbb{R}_+ \times \{0,1\} \times \mathbb{R}^d$ due to

$$J((T_i,\Delta_i,X_i),(t,r,x)) = \int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}\bar{\mathfrak{K}}((s,xX_i),(s',x'))dm_{t',r'}(s')\right)dM_i(t)$$

is a zero-mean $(\mathcal{F}_t)$-martingale. Then, by using the classical theory of V-statistics [28], we obtain

$$n\Psi^2 \overset{\mathcal{D}}{\to} E(J((T_i,\Delta_i,X_i),(T_i,\Delta_i,X_i))) + \mathcal{Y}, \tag{31}$$

where $\mathcal{Y} = \sum_{i=1}^n \lambda_i(\xi_i^2 - 1)$, $\xi_1,\xi_2,\dots$ are independent standard normal random variables, and $\lambda_1,\lambda_2,\dots$ are the eigenvalues associated to the integral operator of J, T^J, defined in Equation (19).

Finally,

$$\begin{aligned}
&\mathbb{E}(J((T_i,\Delta_i,X_i),(T_i,\Delta_i,X_i))) \\
&= \mathbb{E}\left(\int_0^{T_i}\int_0^{T_i}\mathbb{1}_{\{t\neq t'\}}\bar{\mathfrak{K}}((t,X_i),(t',X_i))dM_i(t)dM_i(t')\right) + \mathbb{E}\left(\bar{\mathfrak{K}}((T_i,X_i),(T_i,X_i))\Delta_i\right) \\
&= \int_{x\in\mathbb{R}^d}\int_0^\tau \bar{\mathfrak{K}}((t,x),(t,x))S_{C|X=x}(t)dF_Z(t)dF_X(x)
\end{aligned}$$

where the expectation of the first term is equal to zero by the double martingale Theorem B.7. ■

D.2.4 Proof of Theorem 4.5

Proof: Let $Q : (\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^d)^2 \to \mathbb{R}$ be given by

$$Q((z_1, c_1, x_1), (z_2, c_2, x_2)) = \mathbb{1}_{\{z_1 \leq c_1\}} \mathbb{1}_{\{z_2 \leq c_2\}} \bar{\mathfrak{K}}((z_1, x_1), (z_2, x_2)).$$

Notice that $\Delta_i \Delta_j \bar{\mathfrak{K}}((T_i, X_i), (T_j, X_j)) = Q((Z_i, C_i, X_i), (Z_j, C_j, X_j))$, and thus the asymptotic distribution of $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \Delta_i \Delta_j \bar{\mathfrak{K}}((T_i, X_i), (T_j, X_j))$ can be found by studying the asymptotic distribution of the degenerate V-statistic,

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n Q((Z_i, C_i, X_i), (Z_j, C_j, X_j)).$$

Let $Q_{ij} = Q((Z_i, C_i, X_i), (Z_j, C_j, X_j))$, and notice that $\mathbb{E}(|Q_{ii}|) < \infty$ and $\mathbb{E}(Q_{ij}^2) < \infty$ for $i \neq j$, since, by Assumption 4.2, the kernel $\mathfrak{K}$ is bounded. Then, by the classical theory of U/V-statistics [28],

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n Q_{ij} \xrightarrow{\mathcal{D}} \int_{\mathbb{R}^d} \int_0^\tau \bar{\mathfrak{K}}((t, x), (t, x)) S_{C|X=x}(t) dF_Z(t) dF_X(t) + \mathcal{Y}',$$

where $\mathcal{Y}' = \sum_{i=1}^\infty \lambda_i'(\xi_i^2 - 1)$, $\xi_1, \xi_2, \ldots$ are independent standard normal random variables and $\lambda_1,', \lambda_2', \ldots$ are the eigenvalues associated to the integral operator of Q, T^Q. By Theorem C.5, T^J (the integral operator of J) and T^Q share the same set of non-zero eigenvalues (including multiplicities), and thus the result holds. ■

D.2.5 Proof of lemma 4.6

Proof: Observe that

$$\|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 = V_{0,0} - 2V_{0,1} + V_{1,1},$$

where

$$V_{0,0} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n \mathfrak{K}((T_i, X_j), (T_l, X_k)) \frac{Y_j(T_i)}{Y(T_i)} \frac{Y_k(T_l)}{Y(T_l)} \Delta_i \Delta_l \tag{32}$$

$$V_{0,1} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \mathfrak{K}((T_i, X_j), (T_l, X_l)) \Delta_i \Delta_l \frac{Y_j(T_i)}{Y(T_i)} \tag{33}$$

$$V_{1,1} = \frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \mathfrak{K}((T_i, X_i), (T_l, X_l)) \Delta_i \Delta_l, \tag{34}$$

holds by the reproducing kernel property. The desired result $\|\phi_0^n - \phi_1^n\|_{\mathcal{H}}^2 \xrightarrow{\mathbb{P}} \|\phi_0 - \phi_1\|_{\mathcal{H}}^2$ follows by proving

i) $V_{0,0} \xrightarrow{\mathbb{P}} \|\phi_0\|_{\mathcal{H}}^2$,

ii) $V_{0,1} \xrightarrow{\mathbb{P}} \langle \phi_0, \phi_1 \rangle_{\mathcal{H}}$,

iii) $V_{1,1} \xrightarrow{\mathbb{P}} \|\phi_1\|_{\mathcal{H}}^2$.

We start proving iii). Note that $V_{1,1}$ is a V-statistic of order 2, and note that, by Assumption 4.2, $\mathbb{E}(|\mathfrak{K}((T_i, X_i), (T_i, X_i))\Delta_i|) < \infty$ and $\mathbb{E}(|\mathfrak{K}((T_i, X_i), (T_j, X_j))\Delta_i \Delta_j|) < \infty$ for $i \neq j$. Then, by the classical theory of V-statistics, it holds

$$\begin{aligned}
V_{1,1} &\xrightarrow{\mathbb{P}} \mathbb{E}(\mathfrak{K}((T_1, X_1), (T_2, X_2))\Delta_1 \Delta_2) \\
&= \int_0^\tau \int_{x \in \mathbb{R}^d} \int_0^\tau \int_{x' \in \mathbb{R}^d} \mathfrak{K}((t, x), (t', x')) S_{C|X=x}(t) S_{C|X=x'}(t') dF_{ZX}(t', x') dF_{ZX}(t, x) \\
&= \|\phi_1\|_{\mathcal{H}}^2.
\end{aligned}$$

We continue proving i) and ii). Since the proofs of $i)$ and $ii)$ use the same arguments we just prove the result for $i)$. Rewrite $V_{0,0}$ as

$$V_{0,0} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n f(i, j, l, k)$$

where

$$f(i,j,l,k) = \mathfrak{K}((T_i, X_j), (T_l, X_k))\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)}\Delta_i\Delta_l,$$

and notice that

$$f(i,j,l,k) \leq C\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)}$$

for some constant $C \geq 0$ due to Asssumption 4.2.

We start proving that the sum over elements in $V_{0,0}$ that consider at least the repetition of one index (i,j,l,k) converge to zero. Define $\mathcal{S} = \{(i,j,l,k) \in [n]^4 : |\{i,j,l,k\}| \leq 3\}$, e.g. $(i,i,l,k) \in \mathcal{S}$, and let $\mathcal{S}_{i=l}$ be the set containing all the indices (i,j,l,k) such that $i = l$. It is not hard to see that $\mathcal{S} \subseteq \mathcal{S}_{i=l} \cup \mathcal{S}_{i=j} \cup \mathcal{S}_{i=k} \cup \mathcal{S}_{l=k} \cup \mathcal{S}_{l=j} \cup \mathcal{S}_{j=k}$ (note that the intersection of these sets may not be empty).

Notice that the sum over $\mathcal{S}_{i=l}$ satisfies

$$\frac{1}{n^2}\sum_{\mathcal{S}_{i=l}} f(i,j,l,k) \leq \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n C\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_k(T_i)}{Y(T_i)} \leq \frac{1}{n^2}\sum_{i=1}^n C \to 0.$$

as n tends to infinity.

The sum over $\mathcal{S}_{i=l}$ (or, by symmetry, over $\mathcal{S}_{l=k}$) satisfies

$$\begin{aligned}\frac{1}{n^2}\sum_{\mathcal{S}_{i=j}} f(i,j,l,k) &\leq \frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n\sum_{k=1}^n C\frac{Y_i(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)} \leq \frac{C}{n}\sum_{i=1}^n\frac{1}{Y(T_i)}\\ &\leq \frac{C}{n}\sum_{i=1}^n\frac{1}{n-i+1} \leq \frac{C}{n}\sum_{k=1}^n\frac{1}{k} \leq C\frac{\log(n)+1}{n} \to 0,\end{aligned}$$

as n grows to infinity.

The sum over $\mathcal{S}_{i=k}$ (or, by symmetry, over $\mathcal{S}_{l=j}$) satisfies

$$\frac{1}{n^2}\sum_{\mathcal{S}_{i=k}} f(i,j,l,k) \leq \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n C\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_i(T_l)}{Y(T_l)} \leq \frac{C}{n}\sum_{l=1}^n\frac{1}{Y(T_l)} \to 0.$$

Finally, the sum over $\mathcal{S}_{j=k}$, satisfies

$$\begin{aligned}\frac{1}{n^2}\sum_{\mathcal{S}_{j=k}} f(i,j,l,k) &\leq \frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n\sum_{k=1}^n C\frac{Y_k(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)} \leq \frac{C}{n}\sum_{l=1}^n\frac{1}{Y(T_l)}\\ &\leq \frac{C}{n}\sum_{l=1}^n\frac{1}{n-l+1} \leq \frac{C}{n}\sum_{k=1}^n\frac{1}{k} \leq C\frac{\log(n)+1}{n} \to 0,\end{aligned}$$

as n grows to infinity.

The previous results imply

$$V_{0,0} = \frac{1}{n^2}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i, X_j), (T_l, X_k))\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)}\Delta_i\Delta_l + o(1),$$

where $\mathcal{S}^c = \{(i,j,l,k) \in [n]^4 : |\{i,j,l,k\}| = 4\}$.

We continue by assuming that $Y(t)/n$ can be replaced by its limit, $S_T(t)$, in the previous equation. By doing this, we obtain

$$V_{0,0} = \frac{1}{n^4}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i, X_j), (T_l, X_k))\frac{Y_j(T_i)}{S_T(T_i)}\frac{Y_k(T_l)}{S_T(T_l)}\Delta_i\Delta_l + o_p(1), \tag{35}$$

which is U-statistics of order 4. Then, by using the law of large numbers for U-statistics [28], it follows that

$$\begin{aligned}V_{0,0} &= \mathbb{E}\left(\mathfrak{K}((T_1, X_3), (T_2, X_4))\frac{Y_3(T_1)}{S_T(T_1)}\frac{Y_4(T_2)}{S_T(T_2)}\Delta_1\Delta_2\right) + o_p(1)\\ &= \int_0^\tau\int_{x\in\mathbb{R}^d}\int_0^\tau\int_{x'\in\mathbb{R}^d} K((t,x),(t',x'))d\nu_0(t',x')d\nu_0(t,x) + o_p(1),\\ &= \|\phi_0\|^2_{\mathcal{H}} + o_p(1)\end{aligned}$$

from which we conclude the desired result.

We finalise our proof by proving the replacement of $Y(t)$ by $nS_T(t)$ made in equation (35). By the triangular inequality, it holds

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i,X_j),(T_l,X_k))\left(\frac{Y_j(T_i)}{Y(T_i)}\frac{Y_k(T_l)}{Y(T_l)}-\frac{Y_j(T_i)}{nS_T(T_i)}\frac{Y_k(T_l)}{nS_T(T_l)}\right)\Delta_i\Delta_l\right| \\
&\quad\leq\left|\frac{1}{n^2}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i,X_j),(T_l,X_k))\frac{Y_j(T_i)}{Y(T_i)}\left(\frac{Y_k(T_l)}{Y(T_l)}-\frac{Y_k(T_l)}{nS_T(T_l)}\right)\Delta_i\Delta_l\right| \quad (36)\\
&\quad\quad+\left|\frac{1}{n^2}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i,X_j),(T_l,X_k))\frac{Y_k(T_l)}{S_T(T_l)}\left(\frac{Y_j(T_i)}{Y(T_i)}-\frac{Y_j(T_i)}{nS_T(T_i)}\right)\Delta_i\Delta_l\right|. \quad (37)
\end{aligned}$$

We prove that equations (36) and (37) are both $o_p(1)$ under Assumption 4.2. Since the proofs use the same arguments, we only show the result for equation (36).

Under Assumption 4.2 (bounded $\mathfrak{K}$ for some $C>0$), it holds

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{\mathcal{S}^c}\mathfrak{K}((T_i,X_j),(T_l,X_k))\frac{Y_j(T_i)}{Y(T_i)}\left(\frac{Y_k(T_l)}{Y(T_l)}-\frac{Y_k(T_l)}{nS_T(T_l)}\right)\Delta_i\Delta_l\right| \\
&\quad\leq\frac{C}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{k=1}^n\frac{Y_j(T_i)}{Y(T_i)}\left|\frac{Y_k(T_l)}{Y(T_l)}-\frac{Y_k(T_l)}{nS_T(T_l)}\right|\Delta_l \\
&\quad\leq\frac{C}{n}\sum_{l=1}^n\sum_{k=1}^n\left|\frac{Y_k(T_l)}{Y(T_l)}-\frac{Y_k(T_l)}{nS_T(T_l)}\right|\Delta_l \\
&\quad\leq\frac{C}{n}\int_0^\tau\sum_{k=1}^n\frac{Y_k(t)}{Y(t)}\left|1-\frac{Y(t)/n}{S_T(t)}\right|dN(t) \\
&\quad\leq\frac{C}{n}\int_0^\tau\left|1-\frac{Y(t)/n}{S_T(t)}\right|dN(t)=o_p(1),
\end{aligned}$$

where the last equality follows from Lemma C.6. ■

D.2.6 Proof of Theorem 4.7

Proof: From Lemma 4.6, we have $\Psi_n^2 \xrightarrow{\mathbb{P}} \|\phi_1-\phi_0\|_{\mathcal{H}}^2$ under Assumption 4.2. We need to check then that $\|\phi_1-\phi_0\|_{\mathcal{H}}^2>0$, where $\phi_i=\int_0^\tau\int_{x\in\mathbb{R}^d}K(\cdot,x)L(\cdot,t)d\nu_i(t,x)$ for $i\in\{0,1\}$. By [17, Proposition 1], under the conditions assumed on the kernels, both K and L are c_0-universal, which means that the embedding finite signed measures are injective. Notice, however, that this does not mean that the product kernel is c_0-universal. Nevertheless, by replicating the proof of [17, Theorem 2] (in which we take $\theta=\nu_0-\nu_1$), we can show that $\|\phi_1-\phi_0\|=0$ if and only if $\nu_0=\nu_1$. This result, combined with Proposition 3.3, which shows that $\nu_0=\nu_1$ if and only if $Z\perp X$ proves that $\|\phi_1-\phi_0\|_{\mathcal{H}}^2>0$ under the alternative hypothesis. ■

D.3 Proofs of Section 5

D.3.1 Proof of Lemma 5.1

Proof: The desired result is obtained by showing that

$$\sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}\leq 1}\frac{1}{n}\sum_{i=1}^n\int_0^\tau W_i\left(\bar{\omega}_n(t)-\bar{\omega}(t)\right)dN_i(t)=o_p(n^{-1/2}),$$

where $\bar{\omega}(t)$, given by $\bar{\omega}(t)=\int_{\mathbb{R}^d}\omega(t,x)\frac{S_{T|X=x}(t)}{S_T(t)}dF_X(x)$, is the population version of $\bar{\omega}_n(t)=\sum_{j=1}^n\omega(t,X_j)\frac{Y_j(t)}{Y(t)}$.

Let $\omega^\star(t)=\frac{1}{nS_T(t)}\sum_{j=1}^n\omega(t,X_j)Y_j(t)$. By the triangular inequality,

$$\sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}\leq 1}\frac{1}{n}\sum_{i=1}^n\int_0^\tau W_i\left(\bar{\omega}_n(t)-\bar{\omega}(t)\right)dN_i(t)\leq E_1+E_2,$$

where

$$
\begin{aligned}
E_1 &= \sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}\leq 1} \frac{1}{n}\sum_{i=1}^n \int_0^\tau W_i\left(\bar{\omega}_n(t)-\omega^\star(t)\right)dN_i(t)\\
&= \sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}^2\leq 1} \frac{1}{n}\sum_{i=1}^n \int_0^\tau W_i \sum_{j=1}^n \omega(t,X_j)Y_j(t)R_n(t)dN_i(t), \qquad (38)
\end{aligned}
$$

and

$$
\begin{aligned}
E_2 &= \sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}\leq 1} \frac{1}{n}\sum_{i=1}^n \int_0^\tau W_i\left(\bar{\omega}_n(t)-\omega^\star(t)\right)dN_i(t)\\
&= \sup_{\omega\in\mathcal{H},\|\omega\|^2\leq 1} \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{W_i}{nS_T(t)}\left(\sum_{j=1}^n \omega(t,X_j)Y_j(t)-g(t)\right)dN_i(t), \qquad (39)
\end{aligned}
$$

where $R_n(t)=\frac{1}{Y(t)}-\frac{1}{nS_T(t)}$, and $g(t)=\int_{\mathbb{R}^d}\omega(t,x)S_{T|X=x}(t)dF_X(x)$. Then, the result follows by proving that both, E_1 and E_2, are $o_p(n^{-1/2})$.

Let $\beta\in(0,1)$, and define

$$
B_{n,\beta}=\left\{\frac{Y(t)/n}{S_T(t)}\leq\beta^{-1},\quad \forall t\leq\tau_n\right\}, \qquad (40)
$$

which, by Proposition B.2.1, satisfies $\mathbb{P}(B_{n,\beta})\geq 1-\beta$. We next prove that $n^{1/2}E_1$ converges to zero in probability in the set $B_{n,\beta}$, which implies convergence in probability in the whole space, as $\beta>0$ can be chose arbitrarily small.

Observe

$$
\begin{aligned}
nE_1^2 &= \left(\sup_{\omega\in\mathcal{H}:\|\omega\|_{\mathcal{H}}^2\leq 1}\frac{1}{n^{1/2}}\sum_{i=1}^n\int_0^\tau W_i\sum_{j=1}^n\omega(t,X_j)Y_j(t)R_n(t)dN_i(t)\right)^2\\
&= \frac{1}{n}\sum_{i,l=1}^n W_iW_l\int_0^\tau\int_0^\tau\sum_{j,k=1}^n \mathfrak{K}((t,X_j),(s,X_k))Y_j(t)Y_k(s)R_n(t)R_n(s)dN_i(t)dN_l(s)\\
&\leq \frac{C}{n}\sum_{i,l=1}^n W_iW_l\int_0^\tau\int_0^\tau Y(t)Y(s)|R_n(t)||R_n(s)|dN_i(t)dN_l(s),
\end{aligned}
$$

where the first equality is due to the fact that we are taking supremum in the unit ball of a RKHS, and the inequality in the second line follows from Assumption 4.2 (the kernel is bounded by some constant $C>0$).

Additionally, notice that

$$
\mathbb{E}(nE_1^2\mathbb{1}_{\{B_{n,\beta}\}})\leq\mathbb{E}\left(\frac{C}{n}\sum_{i=1}^n\left(\int_0^\tau Y(t)|R_n(t)|dN_i(t)\right)^2\mathbb{1}_{\{B_{n,\beta}\}}\right), \qquad (41)
$$

since $W_1,\ldots,W_n$ are i.i.d. Rademacher random variables which are independent of our data $((T_i,\Delta_i,X_i))_{i=1}^n$.

We continue by proving that the expectation in the right-hand side of Equation (41) converges to zero. We prove this result by using dominated convergence, and thus we proceed to check its conditions. First, notice that $Y(t)|R_n(t)|=\left|1-\frac{Y(t)/n}{S_T(t)}\right|\leq 1+\beta^{-1}$ uniformly for all $t\leq\tau_n$ in the set $B_{n,\beta}$, and thus

$$
\frac{C}{n}\sum_{i=1}^n\left(\int_0^\tau Y(t)|R_n(t)|dN_i(t)\right)^2\mathbb{1}_{\{B_{n,\beta}\}}\leq C(1+\beta^{-1})^2. \qquad (42)
$$

Second, notice that

$$
\frac{C}{n}\sum_{i=1}^n\left(\int_0^\tau Y(t)|R_n(t)|dN_i(t)\right)^2=\frac{C}{n}\int_0^\tau\left|1-\frac{Y(t)/n}{S_T(t)}\right|^2dN(t)=o_p(1),
$$

by Lemma C.6. Then, by the dominated convergence theorem, and by Equation (41), we deduce $\mathbb{E}(nE_1^2\mathbb{1}_{\{B_{n,\beta}\}})\to 0$, which deduces the desired result $E_1=o_p(n^{-1/2})$.

We continue by proving that $E_2 = o_p(n^{-1/2})$. Notice that it is enough to prove the result in the set $B_{n,\beta}$ defined in Equation (40), as β can be chosen arbitrarily small.

Observe that

$$\begin{aligned}
nE_2^2 &= \left(\sup_{\omega\in\mathcal{H}, \|\omega\|^2\leq 1} \frac{1}{n^{1/2}} \sum_{i=1}^n \int_0^\tau \frac{W_i}{nS_T(t)} \left(\sum_{j=1}^n \omega(t, X_j)Y_j(t) - g(t) \right) dN_i(t) \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{l=1}^n \int_0^\tau \int_0^\tau \frac{W_i W_l}{n^2 S_T(t)S_T(s)} \mathfrak{K}^\star(t,s) dN_i(t) dN_l(s),
\end{aligned}$$

where

$$\begin{aligned}
\mathfrak{K}^\star(t,s) = \sum_{j=1}^n \sum_{k=1}^n \Bigg(&\mathfrak{K}((t,X_j),(s,X_k))Y_j(t)Y_k(s) - \int_{\mathbb{R}^d} \mathfrak{K}((t,X_j),(s,x'))S_{T|X=x'}(s)dF_X(s)Y_j(t) \\
&- \int_{\mathbb{R}^d} \mathfrak{K}((t,x),(s,X_k))S_{T|X=x}(t)dF_X(x)Y_k(s) \\
&+ \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathfrak{K}((t,x),(s,x'))S_{T|X=x}(t)S_{T|X=x'}(s)dF_X(x)dF_X(x') \Bigg),
\end{aligned}$$

follows form the fact we are taking supremum in the unit ball of an RKHS, and from the definition of $g(t)$.

Also, notice that

$$\begin{aligned}
\mathbb{E}(nE_2^2 \mathbb{1}\{B_{n,\beta}\}) &= \mathbb{E}\left(\frac{1}{n^3} \sum_{i=1}^n \int_0^\tau \int_0^\tau \frac{1}{S_T(t)S_T(s)} \mathfrak{K}^\star(t,s) dN_i(t) dN_i(s) \mathbb{1}_{\{B_{n,\beta}\}} \right) \\
&= \mathbb{E}\left(\frac{1}{n^3} \sum_{i=1}^n \int_0^\tau \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} \right) \qquad (43)
\end{aligned}$$

where the first line follows by noticing that $W_1, \dots, W_n$ are i.i.d. Rademacher random variables independent from the data, and the second line follows from the continuity of the survival times T_i.

We use the dominated convergence theorem to show that the expectation in Equation (43) converges to 0. First, notice that by Assumption 4.2,

$$|\mathfrak{K}^\star(t,t)| \leq C\left(Y(t)^2 + 2nY(t)S_T(t) + n^2 S_T(t)^2\right) = C(Y(t) + nS_T(t))^2, \qquad (44)$$

for some constant $C > 0$. Thus,

$$\begin{aligned}
\frac{1}{n^3} \sum_{i=1}^n \int_0^\tau \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} &\leq \frac{C}{n^3} \sum_{i=1}^n \int_0^\tau \left(\frac{Y(t)^2}{S_T(t)^2} + \frac{2nY(t)}{S_T(t)} + n^2 \right) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} \\
&= \frac{C}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{Y(t)/n}{S_T(t)} + 1 \right)^2 dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} \\
&\leq C(\beta^{-1} + 1)^2
\end{aligned}$$

Finally, we need to check that

$$\frac{1}{n^3} \sum_{i=1}^n \int_0^\tau \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} = o_p(1). \qquad (45)$$

Let $\epsilon > 0$, and let $t_\epsilon > 0$ be such that $S_T(t_\epsilon) = \epsilon$. Observe that

$$\begin{aligned}
&\frac{1}{n^3} \sum_{i=1}^n \int_0^\tau \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} \\
&= \frac{1}{n^3} \sum_{i=1}^n \int_0^{t_\epsilon} \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}} + \frac{1}{n^3} \sum_{i=1}^n \int_{t_\epsilon}^\tau \frac{1}{S_T(t)^2} \mathfrak{K}^\star(t,t) dN_i(t) \mathbb{1}_{\{B_{n,\beta}\}}.
\end{aligned}$$

For the first integral, we have

$$\frac{1}{n^3}\sum_{i=1}^n\int_0^{t_\epsilon}\frac{1}{S_T(t)^2}\mathfrak{K}^\star(t,t)dN_i(t)\mathbb{1}_{\{B_n\}} \leq \frac{1}{\epsilon^2}\frac{1}{n^3}\sum_{i=1}^n\int_0^{t_\epsilon}\mathfrak{K}^\star(t,t)dN_i(t)\mathbb{1}_{\{B_n\}}$$
$$= \frac{1}{\epsilon^2}\frac{1}{n^3}\sum_{i=1}^n\mathfrak{K}^\star(T_i,T_i),$$

where the first equality follows from noticing that

$$\mathfrak{K}^\star(t,t) = \left(\sup_{\omega\in\mathcal{H}:\|\omega\|^2_{\mathcal{H}}\leq 1}\sum_{j=1}^n\left(\omega(t,X_j)-\int_{\mathbb{R}^d}\omega(t,x)S_{T|X=x}(t)dF_X(x)\right)\right)^2\geq 0,$$

and by the definition of t_ϵ. Notice that $\frac{1}{n^3}\sum_{i=1}^n\mathfrak{K}^\star(T_i,T_i)$ is a V-statistic of order 3 whose diagonal part converges to zero, and its off-diagonal part has zero mean. Then, by the law of large numbers of V-statistics, we conclude

$$\frac{1}{n^3}\sum_{i=1}^n\int_0^{t_\epsilon}\frac{1}{S_T(t)^2}\mathfrak{K}^\star(t,t)dN_i(t)\mathbb{1}_{\{B_n\}} = o_p(1).$$

For the second integral, by Equation (44) and the definition of the set $B_{n,\beta}$, observe

$$\frac{1}{n^3}\sum_{i=1}^n\int_{t_\epsilon}^{\tau}\frac{1}{S_T(t)^2}\mathfrak{K}^\star(t,t)dN_i(t)\mathbb{1}_{\{B_{n,\beta}\}} \leq \frac{C}{n}\sum_{i=1}^n\int_{t_\epsilon}^{\tau}\left(\frac{Y(t)/n}{S_T(t)}+1\right)^2 dN_i(t)\mathbb{1}_{\{B_{n,\beta}\}}$$
$$\leq C(\beta^{-1}+1)^2\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{T_i\geq t_\epsilon\}}$$
$$= O_p(1)C(\beta^{-1}+1)^2\epsilon,$$

where the last equality follows from the definition of t_ϵ. Since $\epsilon>0$ can be chosen arbitrarily small, we conclude the second integral converges to zero in probability. Thus, by the dominated convergence theorem, we conclude the desired result $E_2 = o_p(n^{-1/2})$.

■

D.3.2 Proof of Proposition 5.4

Proof: As to the computational cost of $n\Psi_n^2$, note first that

$$\text{trace}(\boldsymbol{L}^\Delta(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{K}(\boldsymbol{I}-\boldsymbol{A})^\intercal) = \sum_{i,j=1}^n[\{\boldsymbol{L}^\Delta(\boldsymbol{I}-\boldsymbol{A})\}\circ\{(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{K}]_{ij}.$$

where $\circ$ denotes the elementswise product. The matrix $\boldsymbol{AK}$ can be computed in $\mathcal{O}(n^2)$ because, after sorting the data, $\boldsymbol{A}$ is upper triangular, with the nonzero entries equal along each row. The same observation allows for fast computation of the matrix products in $n\Psi_n^2$ and $(n\Psi_n^2)^W$. ■

D.3.3 Proof of Theorem 5.3

Proof: Given any $\gamma>0$

$$\mathbb{P}(n\Psi_n^2 < Q^W_{n,M}) = \mathbb{P}(n\Psi_n^2 < Q^W_{n,M}, n\Psi_n^2\leq\gamma) + \mathbb{P}(n\Psi_n^2 < Q^W_{n,M}, n\Psi_n^2 > \gamma)$$
$$\leq \mathbb{P}(n\Psi_n^2\leq\gamma) + \mathbb{P}(\gamma < Q^W_{n,M})$$

The first term is dealt by Theorem 4.7 from which we deduce $\limsup_{n\to\infty}\mathbb{P}(n\Psi_n^2\leq\gamma)=0$ (since $n\Psi_n^2\to\infty$). For the second term, observe

$$\mathbb{P}(\gamma < Q^W_{n,M}) \leq \mathbb{P}\left(\bigcup_{m\in\{1,\dots,M\}}\{\gamma < (n(\Psi_n^W)^2)_m\}\right) \leq M\mathbb{P}\left(\gamma < n(\Psi_n^W)^2\right),$$

and notice

$$\begin{aligned}
\limsup_{n\to\infty}\mathbb{P}\left(\gamma<n(\Psi_n^W)^2\right)&=\limsup_{n\to\infty}\mathbb{E}\left(\mathbb{P}\left(\gamma<n(\Psi_n^W)^2|(Z_i,\Delta_i,X_i)_{i=1}^n\right)\right)\\
&\leq\mathbb{E}\left(\limsup_{n\to\infty}\mathbb{P}\left(\gamma<n(\Psi_n^W)^2|(Z_i,\Delta_i,X_i)_{i=1}^n\right)\right)\\
&\leq\mathbb{E}(1-H(\gamma))\\
&=1-H(\gamma),
\end{aligned}$$

where the first inequality follows from the Fatou's lemma, and the second inequality is due to [6, Theorem 3.1], which proves $\limsup_{n\to\infty}\mathbb{P}\left(\gamma<n(\Psi_n^W)^2|(Z_i,\Delta_i,X_i)_{i=1}^n\right)=1-H(\gamma)$, for almost all $(Z_i,\Delta_i,X_i)_{i=1}^\infty$ where H is a distribution function.

Let $\delta>0$, and choose $\gamma>0$ such that $1-H(\gamma)\leq\delta/M$. Then,

$$\limsup_{n\to\infty}\mathbb{P}(\gamma<Q_{n,M}^W)\leq\limsup_{n\to\infty}M\mathbb{P}\left(\gamma<n(\Psi_n^W)^2\right)\leq\delta.$$

Since δ can be chosen arbitrarily small, and since $\limsup_{n\to\infty}\mathbb{P}(n\Psi_n^2\leq\gamma)=0$ for any $\gamma>0$, the result holds ■

E Proofs auxiliary results

E.1 Proof of Proposition C.1

Proof: Recall $\tau_x=\sup\{t:S_{T|X=x}(t)>0\}$ and $\tau=\sup\{t:S_T(t)>0\}$. The next proposition states $\tau_x\leq\tau$ for almost all $x\in\mathbb{R}^d$. We start by proving $\tau'\leq\tau$. Assume otherwise, $\tau<\tau'$, then there exists $\epsilon>0$ such that $B=\{x:\tau_x>\tau+\epsilon\}$ and $F_X(B)>0$. Let

$$B_{1/n}=\left\{x:S_{T|X=x}(\tau+\epsilon)>\frac{1}{n}\right\}$$

and observe $B=\bigcup_{n\geq1}B_{1/n}$. Since $F_X(B)>0$, by union bound, we deduce that there exists $n\geq1$ such that $F_X(B_{1/n})>0$. Then

$$0=\int_{\mathbb{R}^d}S_{T|X=x}(\tau+\epsilon)dF_X(x)\geq\int_{B_{1/n}}\frac{1}{n}dF_X(x)\geq\frac{1}{n}F_X(B_{1/n})>0,$$

from which we deduce that $\tau'\leq\tau$.

We continue proving $\tau'\geq\tau$. Assume that $\tau'<\tau$, then

$$0<S_T(\tau')=\int_{\mathbb{R}^d}S_{T|X=x}(\tau')dF_X(x)=0,$$

from which we deduce $\tau\leq\tau'$, and thus $\tau'=\tau$. ■

E.2 Proof of Proposition C.2

Proof: Recall $d\nu_0^n(s,x)=\frac{1}{n}\sum_{j=1}^n\sum_{i=1}^n\Delta_j\delta_{T_j}(s)\frac{Y_i(s)}{Y(s)}\delta_{X_i}(x)$. Then, the result follows from proving

$$D_n=\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n\Delta_j\omega(T_j,X_i)Y_i(T_j)\left(\frac{n}{Y(T_j)}-\frac{1}{S_T(T_j)}\right)=o_p(1).$$

Notice that D_n can be written as

$$D_n=\frac{1}{n^2}\int_0^{\tau_n}\sum_{i=1}^n\omega(s,X_i)Y_i(s)\left(\frac{n}{Y(s)}-\frac{1}{S_T(s)}\right)dN(s).$$

Let $\epsilon>0$, and define $t_\epsilon\geq0$ such that $S_T(t_\epsilon)=\epsilon$ (notice that $t_\epsilon\geq0$ exists since the distribution over the times is continuous). We decompose D_n into two integrals, one considering integration over $\{s\leq t_\epsilon\}$ and the other one over $\{s>t_\epsilon\}$, and we prove that both integrals converge to zero in probability.

For the first integral, observe that

$$
\begin{aligned}
&\frac{1}{n^2}\int_0^{t_\epsilon}\sum_{i=1}^n \omega(s,X_i)Y_i(s)\left(\frac{n}{Y(s)}-\frac{1}{S_T(s)}\right)dN(s)\\
&\leq \sup_{s\leq t_\epsilon}\left|\frac{n}{Y(s)}-\frac{1}{S_T(s)}\right|\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n|\omega(T_j,X_i)| = o_p(1),
\end{aligned}
$$

where the last equality follows by noticing the supremum converges to zero by Proposition B.1.2., and that $Y_i(s)\leq 1$ and ω are bounded.

For the second integral, it holds

$$
\begin{aligned}
&\frac{1}{n^2}\int_{t_\epsilon}^{\tau_n}\sum_{i=1}^n \omega(s,X_i)Y_i(s)\left(\frac{n}{Y(s)}-\frac{1}{S_T(s)}\right)dN(s)\\
&\leq \frac{C}{n}\int_{t_\epsilon}^{\tau_n}\sum_{i=1}^n\frac{Y_i(s)}{Y(s)}\left|1-\frac{Y(s)/n}{S_T(s)}\right|dN(s)\\
&= O_p(1)\frac{1}{n}\int_{t_\epsilon}^{\tau_n}dN(s) = O_p(1)\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\{t_\epsilon\geq T_j\}} = O_p(1)\epsilon,
\end{aligned}
$$

where the second line follows from the assumption $|\omega|\leq C$ for some constant $C>0$, the first equality in the third line is due to Proposition B.2.1, and the last equality is due to the definition of t_ϵ. Since $\epsilon>0$ can be chosen arbitrarily small, the result holds. ■

E.3 Proof of Lemma C.3

Proof: Notice that $L(t,t)$ is bounded, thus we just need to prove $\sum_{i=1}^n\sum_{j=1}^n\kappa_{ij}(t,t)=o_p(1)$.

Recall $\mathbb{E}'(\cdot)=\mathbb{E}(\cdot|\mathcal{D})$ and the definition of κ_{ij}, then we just need to prove

$$
\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{E}'(K(X_i',X_j)Y_i'(t))Y_j(t)}{nS_T(t)Y(t)},\quad\text{and}\quad \sum_{i=1}^n\sum_{j=1}^n\frac{K(X_i,X_j)Y_i(t)Y_j(t)}{Y(t)^2}
$$

converge to the same limit given by $\mathbb{E}'(K(X_1',X_2')Y_1'(t)Y_2'(t))/S_T(t)^2$. For the first term it holds

$$
\sum_{i=1}^n\sum_{j=1}^n\frac{\mathbb{E}'(K(X_i',X_j)Y_i'(t))Y_j(t)}{nS_T(t)Y(t)} = \frac{1}{n}\sum_{j=1}^n\frac{\mathbb{E}'(K(X_1',X_j)Y_1'(t))Y_j(t)}{S_T(t)Y(t)/n},
$$

and, by the law of large numbers,

$$
\frac{1}{n}\sum_{j=1}^n\frac{\mathbb{E}'(K(X_1',X_j)Y_1'(t))Y_j(t)}{S_T(t)}\overset{a.s.}{\to}\frac{\mathbb{E}(K(X_1,X_2)Y_1(t)Y_2(t))}{S_T(t)},
$$

where $\mathbb{E}'(K(X_1',X_2')Y_1'(t)Y_2'(t))/S_T(t)$ equals $\mathbb{E}(K(X_1,X_2)Y_1(t)Y_2(t))/S_T(t)$. Also, notice that $Y(t)/n\overset{a.s.}{\to}S_T(t)>0$ for any $t<\tau$ Thus, by the continuous mapping theorem, the result holds. The proof for the second term is identical, and thus omitted. ■

E.4 Proof of Lemma C.4

Proof: We start proving the result in *ii*). Under Assumption 4.2, both kernels K and L are bounded, and thus $\left|\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)\right|\leq C$ for some constant $C>0$. Consequently

$$
\left|\frac{1}{n}\int_0^{\tau_n}\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)dN(t)\right|\leq\frac{1}{n}\int_0^{\tau_n}CdN(t)\leq C,
$$

which deduces the desired result in ii).

We continue proving the result in i). Define the process $(W(t))_{t\geq 0}$ by

$$W(t)=\frac{1}{n}\int_0^t\left|\sum_{i=1}^n\sum_{j=1}^n L(s,s)\kappa_{i,j}(s,s)\right|dN(s),$$

and notice that $W(\tau_n)=o_p(1)$ implies i).

Notice that the process $(W(t))_{t\geq 0}$ is a sub-martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$, with compensator (evaluated at $t=\tau_n$) given by

$$A(\tau_n)=\int_0^{\tau_n}\left|\sum_{i=1}^n\sum_{j=1}^n L(s,s)\kappa_{i,j}(s,s)\right|\frac{Y(s)}{n}d\Lambda_Z(s).$$

By the Lenglart-Rebolledo's inequality $A(\tau_n)=o_p(1)$ implies $W(\tau_n)=o_p(1)$. Thus, the result in i) is deduced from $A(\tau_n)=o_p(1)$. We prove $A(\tau_n)=o_p(1)$ by using Dominated Convergence in Probability (DCP) in Lemma B.3. We check the two conditions needed to apply DCP. First, by Lemma C.3,

$$\left|\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)\right|\frac{Y(t)}{n}\leq\left|\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)\right|=o_p(1)$$

point-wise for any $t<\tau$. Second,

$$\left|\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)\right|\frac{Y(t)}{n}=O_p(1)S_Z(t)S_C(t)$$

holds uniformly for all $t\leq\tau_n$, since, by Proposition B.2, $Y(t)/n=S_T(t)$ uniformly for all $t\leq\tau_n$, $S_T(t)=S_Z(t)S_C(t)$ under the null, and since $\left|\sum_{i=1}^n\sum_{j=1}^n L(t,t)\kappa_{i,j}(t,t)\right|$ is bounded by Assumption 4.2. Then, by DCP, we conclude $A(\tau_n)=o_p(1)$, which implies the result in i). ■

E.5 Proof of Theorem C.5

Proof: Consider a re-parametrisation of the kernel J (given in Equation (12) in the main document) in terms of the augmented space (Z,C,X), as $J:(\mathbb{R}_+\times\mathbb{R}_+\times\mathbb{R}^d)^2\to\mathbb{R}$ given by

$$J((Z_1,C_1,X_1),(Z_2,C_2,X_2))=\int_{\mathbb{R}_+}\int_{\mathbb{R}_+}\bar{\mathfrak{K}}((s_1,X_1),(s_2,X_2))dM_1(s_1)dM_2(s_2),$$

where $dM_1(s_1)=\mathbb{1}_{\{Z_1\leq C_1\}}\delta_{Z_1}(s_1)-\mathbb{1}_{\{\min\{Z_1,C_1\}\}\geq s_1}d\Lambda_Z(s_1)$. Notice that to evaluate J in the previous definition we do not need extra information, it only needs (T_1,Δ_1,X_1) and (T_2,Δ_2,X_2) to be evaluated, but it is more convenient theoretically.

From the definition of the backward operator B, and Property B.5.5 (see Section B.5), we deduce

$$J((Z_1,C_1,X_1),(Z_2,C_2,X_2))=(B_1B_2Q)((Z_1,C_1,X_1),(Z_2,C_2,X_2)),\tag{46}$$

where B_1 and B_2 denote the operator B applied to (Z_1,C_1,X_1) and (Z_2,C_2,X_2), respectively; the same definition holds for A_1 and A_2. Thus,

$$(T^Jf)(Z_1,C_1,X_1)=\mathbb{E}_2\left((B_1B_2Q)((Z_1,C_1,X_1),(Z_2,C_2,X_2))f(Z_2,C_2,X_2)\right).$$

By using property B.5.1, and the linearly of B_1, it holds

$$\begin{aligned}(T^Jf)(Z_1,C_1,X_1)&=\mathbb{E}_2\left((B_1Q)((Z_1,C_1,X_1),(Z_2,C_2,X_2))(Af)(Z_2,C_2,X_2)\right)\\&=B\mathbb{E}_2\left(Q((Z_1,C_1,X_1),(Z_2,C_2,X_2))(Af)(Z_2,C_2,X_2)\right)\\&=B(T^Q(Af)),\end{aligned}$$

where T^Q is the integral operator associated to the kernel Q.

Let (λ_i, f_i), with $\lambda_i \neq 0$, be an eigenpair of T^J, then we claim that (λ_i, Af_i) is a eigenpair of T^Q. Indeed,

$$T^Q Af_i = ABT^Q Af_i = AT^J f_i = \lambda_i Af_i,$$

where in the first equality we use property B.5.2. Also, if (λ_j, f_j) is another eigenpair such that $\langle f_i, f_j\rangle_{\mathcal{L}_2} = 0$, then

$$\langle Af_i, Af_j\rangle_{\mathcal{L}_2} = \langle f_i, BAf_j\rangle_{\mathcal{L}_2} = \langle f_i, f_j - \mathbb{E}(f_j|C_1, X_1)\rangle_{\mathcal{L}_2} = 0,$$

since $\mathbb{E}(f_j|C_1, X_1) = \frac{1}{\lambda_j}\mathbb{E}(B(T^Q Af_j)) = 0$ by property B.5.4. Hence, we conclude that for every different eigenpair (λ_i, f_i) belonging to T^J, there exists a corresponding pair (λ_i, Af_i) associated to T^Q.

Conversely for any eigenpair (λ_i', g_i) associated to T^Q, we claim that (λ_i', Bg_i) is a eigenpair for T^J. To see this, observe

$$T^J Bg_i = BT^Q ABg_i = BT^Q g_i = \lambda_i' Bg_i.$$

Similarly, if (λ_j', g_j) is another eigenpair with $\lambda_j \neq 0$, and such that $\langle g_i, g_j\rangle_{\mathcal{L}_2} = 0$, then

$$\langle Bg_i, Bg_j\rangle_{\mathcal{L}_2} = \langle g_i, ABg_j\rangle_{\mathcal{L}_2} = \langle g_i, g_j\rangle_{\mathcal{L}_2} = 0.$$

Hence, we conclude that for every different eigenpair (λ_i', g_i) belonging to T^Q, there exists a corresponding pair (λ_i', Bg_i) associated to T^Q. We conclude that T^Q and T^J have the same set of non-zero eigenvalues including multiplicities. ■

E.6 Proof of Lemma C.6

Proof: We only prove the result for $k = 1$ as the other result follows by using the same arguments.

Let $\epsilon > 0$ and choose $t_\epsilon \geq 0$ such that $S_T(t_\epsilon) = \epsilon$. We split the integral into two integrals, one over $\{t \leq t_\epsilon\}$, and other $\{t > t_\epsilon\}$. For the integral, observe that

$$\frac{1}{n}\int_0^{t_\epsilon}\left|1 - \frac{Y(t)/n}{S_T(t)}\right| dN(t) \leq \sup_{t\leq t_\epsilon}\left|\frac{n}{Y(t)} - \frac{1}{S_T(t)}\right| \frac{1}{n}\int_0^{t_\epsilon} dN(t) = o_p(1),$$

since $\frac{1}{n}\int_0^{t_\epsilon} dN(t) \leq 1$ for all $n \geq 1$, and $\sup_{t\leq t_\epsilon}\left|\frac{n}{Y(t)} - \frac{1}{S_T(t)}\right| \to 0$ by Proposition B.1.2. For the second integral, observe that

$$\begin{aligned}
\frac{1}{n}\int_{t_\epsilon}^{\tau}\left|1 - \frac{Y(t)/n}{S_T(t)}\right| dN(t) &= O_p(1)\frac{1}{n}\int_0^{\tau_n} \mathbb{1}_{\{t>t_\epsilon\}} dN(t)\\
&= O_p(1)\frac{1}{n}\sum_{i=1}^n \Delta_i \mathbb{1}_{\{T_i>t_\epsilon\}}\\
&= O_p(1)\epsilon,
\end{aligned}$$

where the first equality is due to $Y(t)/n = O_p(1)S_T(t)$ uniformly for all $t \leq \tau_n$ by Proposition B.2.1, and the third equality follows from Markov's inequality and the definition of t_ϵ. Since $\epsilon > 0$ is arbitrary, we deduce equation (20). ■