

# Nonparametric Tests for Conditional Treatment Effects with Duration Outcomes \*

Pedro H. C. Sant'Anna †

Job Market Paper

November 5, 2014

## Abstract

This paper proposes new nonparametric tests for treatment effects when the outcome of interest, typically a duration, is subjected to right censoring. Our tests are based on Kaplan-Meier integrals, and do not rely on distributional assumptions, shape restrictions, nor on restricting the potential treatment effect heterogeneity across different subpopulations. The proposed tests are consistent against fixed alternatives and can detect nonparametric alternatives converging to the null at the parametric  $n^{-1/2}$ -rate,  $n$  being the sample size. Finite sample properties of the proposed tests are examined by means of a Monte Carlo study. We illustrate the use of the proposed policy evaluation tools by studying the effect of labor market programs on unemployment duration based on experimental and observational datasets.

**JEL:** C12, C14, C21, C24, C41.

**Keywords:** Censored data; Empirical Process; Kaplan-Meier Integrals; Survival Analysis; Treatment Effects.

---

\*I am most thankful to my advisor Prof. Miguel A. Delgado for his guidance, constant support and encouragement. I am also grateful to Carlos Velasco and Juan Carlos Escanciano for innumerable advices. I also thank Prof. Myoung-jae Lee for providing the Korean data on female job training program, as well as Andrés García-Suaza, Bernhard Schmidpeter, Christos Mavridis, Emre Ergemen, Lian Allub, Mehdi Sahneh, Renata Sant'Anna and seminar participants at Universidad Carlos III, 2014 European Meeting of the Econometric Society, 2014 ZEW Summer Workshop and at ENTER Jamboree 2014 for helpful comments and discussions. Research is funded by Spanish Plan Nacional de I+D+I grant number ECO2012-33053.

†Ph.D Candidate, Universidad Carlos III de Madrid. Contact: pde@eco.uc3m.es. Phone: +34 652096646

# 1 Introduction

Assessing whether a policy has any effect on an outcome of interest has been one of the main concerns in empirical research. As summarized in Imbens (2004), Heckman and Vytlacil (2007), and Imbens and Wooldridge (2009), the focus of the policy evaluation literature has been mainly confined to situations where the realized outcome of interest is completely observed for the treated and the control groups. However, when the outcome variable is subjected to censoring, such inference procedures may provide misleading conclusions on the effect of the proposed policy. Assessing if labor market programs affect the length of unemployment, if correctional programs affect recidivism of criminal activities, or whether the survival time is affected by a new clinical therapy are just few examples where the outcome of interest is usually subjected to censoring mechanisms, and hence, standard policy evaluation procedures are not suitable. This article remedy this by proposing new nonparametric tests for conditional treatment effects when the outcome of interest, typically a duration, is subjected to right censoring.

Our test statistics are suitable functionals of empirical processes whose limiting distributions under the null can be estimated using a multiplicative-type bootstrap, which is proved to be valid. Our proposed tests are consistent against both one and two-sided alternative fixed alternatives and can detect nonparametric alternatives converging to the null at the parametric  $n^{-1/2}$ -rate,  $n$  being the sample size. Since our test proposal does not rely on continuity assumptions regarding the duration outcome, our policy evaluation tools are suitable for both discrete and continuous censored data. Moreover, our tests can be used not only for unconfounded treatment assignments, but also for the local treatment effect setup of Imbens and Angrist (1994) and Angrist et al. (1996), and for the case of dynamic treatment allocations as described in Sianesi (2004). Overall, this paper offers a unifying approach to derive uniformly valid nonparametric tests for treatment effects with censored outcomes. Although our focus is on hypotheses testing, estimators for unconditional treatment effects naturally arises as a by-product of the testing procedure.

To achieve the aforementioned properties, this paper relies on three components. First,

our tests are based on inverse probability weighting (IPW) estimators of the relevant treatment effect measures, in which the propensity score is estimated by nonparametric methods. In particular, we consider the series logit estimator proposed by Hirano et al. (2003), but other estimators are possible. Second, because the focus of this paper is on testing for conditional treatment effects, our hypotheses of interest are based on conditional moment restrictions. To avoid the use of smooth estimates, we adopt an integrated moment approach, reducing the conditional moment restrictions to an infinite number of unconditional orthogonality restrictions, as others have adopted in different contexts, see e.g. Delgado (1993), Stute (1997), Stute et al. (1998) and Delgado and González-Manteiga (2001). In a setup without censoring, we would be able to estimate the integrated moments by their empirical analogue. However, this is not feasible when the outcome of interest is subjected to right censoring. To handle this issue, we characterize the integrated moments as Kaplan-Meier (KM) integrals, see e.g. Stute and Wang (1993*a,b*), Stute (1993, 1995, 1996), and Sellero et al. (2005). However, because the treatment effect measures depend on the propensity score, our integrand is unknown, which is in contrast to the literature on KM integrals. To accommodate this issue, we present new results for Kaplan-Meier integrals indexed by unknown, possibly infinite-dimensional nuisance parameters.

This paper is directly connected to the treatment effects literature. For recent reviews of this huge literature, see e.g. Imbens (2004), Heckman and Vytlačil (2007), and Imbens and Wooldridge (2009), among others. In cases where the outcome is subjected to censoring, few estimation procedures have been considered, see e.g. Ham and Lalonde (1996), Eberwein et al. (1997), Hubbard et al. (2000), Abbring and van den Berg (2003), Abbring and van den Berg (2005), Crépon et al. (2009), and Frandsen (2014), among others. Nonetheless, the aforementioned papers have not devoted attention to nonparametric tests. In fact, the literature on nonparametric tests for treatment effects is scarce, Abadie (2002), Crump et al. (2008), Lee and Whang (2009), Delgado and Escanciano (2013), and Hsu (2013) being exceptions when censoring is not an issue. In the presence of censoring, Lee (2009) developed a nonparametric test of the null hypothesis of no distributional treatment effect. However, the “two sample” setup adopted by Lee (2009) greatly differs from ours.

To illustrate the relevance of our new policy evaluation tools, we apply the proposed tests to evaluate labor market programs using two different sets of applications. First, as in Woodbury and Spiegelman (1987), we analyze the Illinois Reemployment Bonus Experiments that was carried out in the 1980's. Then, as in Lee (2009), we use observational female job training data from the Department of Labor in South Korea to test if receiving job training instead of unemployment insurance affects the unemployment duration. With these applications we show that introducing ad hoc parametric assumptions or ignoring treatment effect heterogeneity may lead to spurious conclusions about the policy effectiveness.

The remainder of the paper is organized as follows. We first describe the basic setup and the concentrate on testing the null of zero conditional distributional treatment effects. In Section 3, we derive the asymptotic distribution for the baseline tests and introduce a bootstrap method to approximate their critical values. A Monte Carlo study in Section 4 investigates the finite sample properties of the test proposals. In Section 5, we present some applications of our basic setup, i.e. we consider the null of zero conditional average treatment effects and show that our test procedure is also suitable when treatment allocation is endogenous or dynamic. In Section 6, we apply the policy evaluation tests to different datasets. Finally, we offer concluding remarks and suggest extensions for future research. Mathematical proofs are gathered in an appendix at the end of the article.

## **2 Testing for zero conditional treatment effects with censored outcomes**

### **2.1 Basic setup**

We consider a set of individuals flowing into a state of interest, and the time these individuals spend in that state is our outcome of interest. Upon inflow, an individual is assigned to a treatment or to a control group. The goal of this paper is to assess different hypotheses related to the causal effect of the treatment on the time spent in this state of interest.

Henceforth, all random variables are defined on a common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Let  $D$  be an indicator of participation in the program, i.e.  $D = 1$  if the unit participates in the treatment and  $D = 0$  otherwise. Define  $Y_0$  and  $Y_1$  as the potential outcomes under the control and treatment groups, respectively. Additionally, let  $X \in \mathbb{R}^k$  be vector of pre-treatment variables, and  $\chi_{Y,X} \subseteq \mathbb{R} \times \mathbb{R}^k$  denote the support of  $Y \times X$ .

In this paper, the treatment effect measure of main interest is the conditional distributional treatment effect, that is, the difference between the conditional cumulative distribution function (*CDF*) of the potential outcome under treatment and control:

$$\Upsilon(t, x) = \mathbb{E}[1\{Y_1 \leq t\} - 1\{Y_0 \leq t\} | X = x].$$

Our main focus is on testing the hypothesis that the distributional treatment effect (DTE) is equal to zero for every subpopulation defined by covariates, that is,

$$H_0 : \Upsilon(t, x) = 0 \quad \forall (x, t) \in \mathcal{W}, \tag{1}$$

where  $\mathcal{W} \subseteq \chi_{Y,X}$ . Under the null hypothesis  $H_0$ , the conditional distribution of  $Y$  is not affected by the treatment at  $\mathcal{W}$ , and the alternative hypothesis  $H_1$  is the negation of  $H_0$ .

An important feature of the hypothesis in (1) is its focus on distributional treatment effects, and not only on the average treatment effects. By doing so, one can assess if the treatment has affected any feature of the distribution of the outcome, and not necessarily just the mean. In fact, by looking at the outcome distribution, one is able to perform welfare analysis under mild assumptions about social preferences, see e.g. Abadie (2002). Such analysis would not be possible if the focus were only at average treatment effects.

Another distinguishing characteristic of (1) is its focus on *conditional* treatment effects, and not only on the *unconditional* treatment effects. That is, in this paper we are concerned about the ubiquitous and commonly ignored feature that treatment effects may vary across different subpopulations. Although heterogeneity in the effect of a policy is generally allowed, *unconditional* measures of treatment effects may neglect some important differences in policy evaluations. For instance, a labor market program that does not affect the

unemployment duration for the overall population might still be effective for a subgroup of individuals with specific observable characteristics. As illustrated by Bitler et al. (2006, 2008, 2014) and Crump et al. (2008), being able to assess if the treatment has affected any subpopulation is a crucial element of policy evaluations.

Next, we describe our setup. In order to model the treatment effect, we adopt the potential outcome notation popularized by Rubin (1974). Let  $D$ ,  $Y_0$ ,  $Y_1$  and  $X$  be defined as before, and let  $p(x) \equiv \mathbb{P}(D = 1|X = x)$  be the propensity score, i.e. the conditional probability of receiving treatment. Although our interest is on  $Y_0$  and  $Y_1$ , one can only observe  $Q \equiv DQ_1 + (1 - D)Q_0$ , where  $Q_0 = \min\{Y_0, C_0\}$ ,  $Q_1 = \min\{Y_1, C_1\}$ ,  $C_0$  and  $C_1$  being potential censoring random variables under the control and treatment groups, respectively. Censoring might appear for different reasons such as the end of a follow-up or drop out. In addition to  $Q$ , one also observe the censoring indicator  $\delta \equiv D\delta_1 + (1 - D)\delta_0$ , where, for  $j \in \{0, 1\}$ ,  $\delta_j = 1\{Y_j \leq C_j\}$ .

**Assumption 1**  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^n$  are independent and identically distributed observations of  $(Q, \delta, D, X)$ .

**Assumption 2**  $(Y_0, Y_1, C_0, C_1) \perp\!\!\!\perp D|X$  a.s.

**Assumption 3** For all  $x \in \mathcal{W}$  and some  $\varepsilon > 0$ ,  $\varepsilon \leq p(x) \leq 1 - \varepsilon$ .

**Assumption 4** Assume that

(i)  $(Y_0, Y_1) \perp\!\!\!\perp (C_0, C_1)$

(ii) For  $j \in \{0, 1\}$ ,  $\mathbb{P}(\delta_j = 1|X, Y_j) = \mathbb{P}(\delta_j = 1|Y_j)$ .

**Assumption 5** The distributions of  $Y_j$  and  $C_j$ ,  $j \in \{0, 1\}$ , has no common jumps

Assumptions 1-3 are standard in the treatment effects literature. Assumption 2 was introduced by Rosenbaum and Rubin (1983), and states that, conditional on observables, treatment assignment is independent of potential outcomes and censoring. Assumption 3 states that there is overlap in the covariate distributions. As shown by Khan and Tamer

(2010), Assumption 3 is crucial in determining the convergence rate of inverse probability weighted estimators.

In the absence of censoring, Rosenbaum and Rubin (1983) show that Assumptions 2 and 3 would suffice to identify different treatment effects measures, in particular  $\Upsilon(t, x)$ . Nonetheless, it is important to notice that censoring introduces an additional identification challenge because the probability of censoring is related to potential outcomes, that is, censoring occurs only if  $Y_j > C_j$ ,  $j \in \{0, 1\}$ . Ignoring the censoring problem or analyzing only the uncensored outcomes would therefore introduce another source of confounding. To overcome such issue, Assumption 4 imposes additional structure on the censoring mechanism.

Assumption 4 states that, given the “time of death”  $Y_j$ , the covariates do not provide any further information whether censoring will take place, that is,  $\delta_j$  and  $X$  are conditionally independent given the potential outcome  $Y_j$ . A particular case in which it holds is when  $C_j$  is independent of  $(Y_j, X)$ , as assumed in Honore et al. (2002), Lee and Lee (2005) and Frandsen (2014), for example. Nonetheless, Assumption 4 is more general and allows censoring to depend on the covariates through the potential outcome  $Y_j$ . We notice that similar assumptions have been used in different contexts, see e.g. Chen (2001), Tang et al. (2003), D’Haultfoeuille (2010) and Breunig et al. (2014). An alternative to Assumption 4 is  $(Y_0, Y_1) \perp\!\!\!\perp (C_0, C_1) | X$ . In this case the use of smoothing techniques and trimming procedures are required, see Akritas (1994), González-Manteiga and Cadarso-Suárez (1994), and Iglesias Pérez and González-Manteiga (1999) for examples in different contexts. With Assumption 4, the use of smothers and trimming is avoided.

Assumption 5 is a regularity condition that does not exclude discontinuities of  $F_{Y_j}(\cdot) \equiv \mathbb{P}(Y_j \leq \cdot)$  and  $G_j(\cdot) \equiv \mathbb{P}(C_j \leq \cdot)$  at distinct points, that is, we do not impose that  $F_{Y_j}$  and  $G_j$  must be absolutely continuous. Therefore, we allow for both discrete and continuous potential outcomes.

With the aforementioned assumptions, the next proposition shows that we can point identify  $\Upsilon(t, x)$  from the  $(Q, \delta, D, X)$ . For  $j \in \{0, 1\}$ , let  $\tau_{C_j} = \sup \{t : G_j(t) < 1\}$ . For simplicity, assume that  $\tau_{C_0} = \tau_{C_1} = \tau_C$ .

**Proposition 1** Under Assumptions 2-4, for  $(t, x) \in (-\infty, \tau_C) \times \mathbb{R}^k$ ,

$$\Upsilon(t, x) = \mathbb{E} \left[ \left( \frac{D \delta 1 \{Q \leq t\}}{(1 - G_1(Q-)) p(X)} - \frac{(1 - D) \delta 1 \{Q \leq t\}}{(1 - p(X)) (1 - G_0(Q-))} \right) \middle| X = x \right]. \quad (2)$$

Some remarks are necessary. From Proposition 1, one can see that nonparametric point identification of the distributional treatment effect over the entire outcome support may not be feasible. This is intuitive because, due to right censoring mechanisms, potential outcomes beyond  $\tau_C$  are never observed. Given that one may not point identify the whole distributional treatment effect, the point identification of traditional measures such as the average treatment effect  $\mathbb{E}[Y_1 - Y_0]$  is also at stake<sup>1</sup>. Nonetheless, (2) has considerable identification power. That is, by focusing on  $\mathcal{W} \subseteq (-\infty, \tau_C) \times \mathbb{R}^k$ , one can still point identify the distributional treatment effects measure of interest and test the hypothesis (1) within this portion of the *CDF*. This is feasible because  $\tau_C$  is usually known in applications.

Another important feature of (2) is that the potentially restrictive condition that the censoring distribution is the same under both treatment regimes is not necessary for identification. Such result is in contrast with the one in Frandsen (2014), for example. Indeed, if one assumes that the censoring distribution is the same but this condition is not fulfilled, treatment effects measures may suffer from severe bias and tests based on this assumption may have large size distortions; see Section 4.

## 2.2 Characterization of the null hypothesis

Given that  $\Upsilon(t, x)$  is identified from the data, we are able to characterize the null hypothesis (1) in terms of observables. In fact, based on the representation in (2) and using Assumption 3 guaranteeing that  $p(\cdot) \in (0, 1)$ , we have,

$$\Delta(t, x) = 0 \forall (x, t) \in \mathcal{W} \Leftrightarrow \Upsilon(t, x) = 0 \forall (x, t) \in \mathcal{W}$$

---

1. In Section 5, we show how one can identify a related measure, the trimmed ATE.

where

$$\begin{aligned}\Delta(t, x) &= \mathbb{E} \left[ \left( \frac{D(1-p(X))}{(1-G_1(Q-))} - \frac{(1-D)p(X)}{(1-G_0(Q-))} \right) \delta 1\{Q \leq t\} \middle| X = x \right] \\ &= \Upsilon(t, x) p(x) (1 - p(x)).\end{aligned}$$

That is, in order to test the null hypothesis (1), it suffices to check if  $\Delta = 0$ . The main advantage of focusing on  $\Delta(\cdot, \cdot)$  instead of  $\Upsilon(\cdot, \cdot)$  is that random denominators due to the propensity score are avoided.

In order to assess if  $\Delta(\cdot, \cdot) = 0$ , there are two main approaches. The first one consists of using nonparametric smooth estimates of  $\Delta$ . An important limitation of this local approach arises when  $X$  is multivariate due to the ‘‘curse of dimensionality’’. Moreover, tests in this category are not able to detect local alternatives converging to the null at the parametric rate  $n^{-1/2}$ . Instead, we adopt an integrated moment approach, avoiding the use of smoothers by means of reducing the conditional moment restriction to an infinite number of unconditional orthogonality restrictions, i.e., we characterize the null hypothesis (1) as

$$H_0 : I(t, x) = 0 \quad \forall (t, x) \in \mathcal{W}, \quad (3)$$

where

$$I(t, x) = \mathbb{E} \left[ \left( \frac{D(1-p(X))}{(1-G_1(Q-))} - \frac{(1-D)p(X)}{(1-G_0(Q-))} \right) \delta 1\{Q \leq t\} 1\{X \leq x\} \right]. \quad (4)$$

This integrated approach has been used in different contexts, see e.g. Delgado (1993), Stute (1997), Stute et al. (1998), Koul and Stute (1999) and Delgado and González-Manteiga (2001). Although other characterizations of  $H_0$  are feasible ( see Bierens and Ploberger (1997), Stinchcombe and White (1998) and Escanciano (2006*a,b*)), we do not pursue these possibilities in this paper.

### 2.3 Kaplan-Meier integrals and test statistics

The characterization of the null hypothesis in (3) suggests using functionals of an estimator of  $I(\cdot, \cdot)$  as test statistics. Therefore, we must first estimate  $I(\cdot, \cdot)$  using the sample  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^n$ . From (4), the challenge of estimating  $I(\cdot, \cdot)$  is reduced to estimating  $p(\cdot)$ ,  $G_1(\cdot)$ ,  $G_0(\cdot)$ , and then applying the plug-in principle.

The task of nonparametrically estimate  $p(\cdot)$  is relatively standard. Following Hirano et al. (2003), we can nonparametrically estimate  $p(\cdot)$  using the Series Logit Estimator (SLE) based on power series. Although other nonparametric estimators could be used - see e.g. Ichimura and Linton (2005) and Li et al. (2009) - we do not exploit these possibilities in this paper.

To define the SLE, let  $\lambda = (\lambda_1, \dots, \lambda_r)'$  be a  $r$ -dimensional vector of non-negative integers with norm  $|\lambda| = \sum_{j=1}^r \lambda_j$ . Let  $\{\lambda(l)\}_{l=1}^\infty$  be a sequence including all distinct multi-indices  $\lambda$  such that  $|\lambda(l)|$  is non-decreasing in  $l$  and let  $x^\lambda = \prod_{j=1}^r x_j^{\lambda_j}$ . For any integer  $L$ , define  $R^L(x) = (x^{\lambda(1)}, \dots, x^{\lambda(L)})'$  as a vector of power functions. Let  $\mathcal{L}(a) = \exp(a) / (1 + \exp(a))$  be the logistic *CDF*. The SLE for  $p(x)$  is defined as  $\hat{p}(x) = \mathcal{L}(R^L(x)' \hat{\pi}_L)$ , where

$$\hat{\pi}_L = \arg \max_{\pi_L} \frac{1}{n} \sum_{i=1}^n D_i \log(\mathcal{L}(R^L(X_i)' \pi_L)) + (1 - D_i) \log(1 - \mathcal{L}(R^L(X_i)' \pi_L)).$$

Next, instead of directly considering estimators for  $G_1(\cdot)$  and  $G_0(\cdot)$ , we show that, similarly to Stute (1993, 1996), we can estimate  $I(\cdot, \cdot)$  by means of empirical Kaplan-Meier integrals. To fix ideas, suppose we could fully observe  $(Y, X, D)$ , implying that  $G_1(\cdot) = G_0(\cdot) = 0$  a.s.. For a given  $(t, x) \in \mathcal{W}$ , define

$$\xi_1(\bar{y}, \bar{x}, \bar{z}; t, x) = \bar{z}(1 - p(\bar{x})) 1\{\bar{y} \leq t\} 1\{\bar{x} \leq x\}, \quad (5)$$

$$\xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) = (1 - \bar{z}) p(\bar{x}) 1\{\bar{y} \leq t\} 1\{\bar{x} \leq x\}, \quad (6)$$

and notice that, in the absence of censoring,

$$\begin{aligned} I(t, x) &= \mathbb{E}[\xi_1(Y, X, D; t, x)] - \mathbb{E}[\xi_0(Y, X, D; t, x)] \\ &= \int \xi_1(\bar{y}, \bar{x}, \bar{z}; t, x) F_1(d\bar{y}, d\bar{x}) - \int \xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) F_0(d\bar{y}, d\bar{x}), \end{aligned}$$

where  $F_j(t, x) \equiv \mathbb{P}(Y \leq t, X \leq x, D = j)$ ,  $j \in \{0, 1\}$ .

From the above representation, and with the SLE  $\hat{p}(\cdot)$  at our disposal, one could estimate  $I(\cdot, \cdot)$  by its sample analogue

$$\begin{aligned} &\int \hat{\xi}_1(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}_1(d\bar{y}, d\bar{x}) - \int \hat{\xi}_0(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}_0(d\bar{y}, d\bar{x}) \quad (7) \\ &= \frac{1}{n} \sum_{i=1}^n \left[ \hat{\xi}_1(Y_i, X_i, D_i; t, x) - \hat{\xi}_0(Y_i, X_i, D_i; t, x) \right] \end{aligned}$$

where  $\hat{F}_j(t, x)$  denotes the empirical analog of  $F_j(t, x)$ , and

$$\hat{\xi}_1(\bar{y}, \bar{x}, \bar{z}; t, x) = \bar{z} (1 - \hat{p}(\bar{x})) 1\{\bar{y} \leq t\} 1\{\bar{x} \leq x\}, \quad (8)$$

$$\hat{\xi}_0(\bar{y}, \bar{x}, \bar{z}; t, x) = (1 - \bar{z}) \hat{p}(\bar{x}) 1\{\bar{y} \leq t\} 1\{\bar{x} \leq x\}, \quad (9)$$

the analogous of (5) and (6), but with the true  $p(\cdot)$  replaced by the SLE  $\hat{p}(\cdot)$ . Unfortunately, due to the censoring problem,  $\hat{F}_j(\cdot, \cdot)$  is not at our disposal and therefore, the above procedure is infeasible. Nonetheless, we can exploit other possibilities. Since the Kaplan and Meier (1958) estimator is the analogous to the empirical CDF when the outcome is subjected to right censoring, a convenient way to proceed involves using some multivariate Kaplan-Meier (KM) estimator of  $F_j(\cdot, \cdot)$ , which would use only the information available at the sample  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^n$ .

To define the KM estimator of  $F_j(t, x)$ ,  $j = 0, 1$ , let  $n_1$  and  $n_0$  be the total number of individuals in the treated and control subsamples,  $Q_{j,1:n_j} \leq \dots \leq Q_{j,n_j:n_j}$  be the ordered  $Q$  values for the subsamples with  $D = j \in \{0, 1\}$ , where ties within  $Y$  or within  $C$  are ordered arbitrarily and ties among  $Y$  and  $C$  are treated as if the former precedes the later, and let  $\delta_{j,[i:n_j]}$  and  $X_{j,[i:n_j]}$  be the concomitant of the  $i$ th order statistics of the subsample with

$D = j$ , i.e. the  $\delta$  and  $X$  paired with  $Q_{j,i:n_j}$ . Similarly to Stute (1993, 1996), the multivariate Kaplan-Meier estimator of  $F_j(t, x)$  is given by

$$\hat{F}_j^{KM}(t, x) = \sum_{i=1}^{n_j} W_{j,i:n_j} 1\{Q_{j,i:n_j} \leq t\} 1\{X_{j,[i:n_j]} \leq x\},$$

where

$$W_{j,k:n_j} = \frac{n_j}{n} \frac{\delta_{j,[k:n_j]}}{n_j - k + 1} \prod_{l=1}^{k-1} \left( \frac{n_j - l}{n_j - l + 1} \right)^{\delta_{j,[l:n_j]}}$$

denotes its ‘‘jump’’ at observation  $k$ . It is important to notice that, because we do not impose that the censoring variables  $C_1$  and  $C_0$  follow the same distribution, the KM jump differ depending on whether  $D$  is equal to 0 or 1. This is the reason why we must consider different KM estimators for  $F_0(\cdot, \cdot)$  and  $F_1(\cdot, \cdot)$ .

With the SLE  $\hat{p}(\cdot)$  and the KM estimators  $\hat{F}_1^{KM}(\cdot, \cdot)$  and  $\hat{F}_0^{KM}(\cdot, \cdot)$  at hands, one can follow the same steps as in (7), and estimate  $I(\cdot, \cdot)$  by

$$\begin{aligned} \hat{I}(t, x) &= \int \hat{\xi}_1(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}_1^{KM}(d\bar{y}, d\bar{x}) - \int \hat{\xi}_0(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}_0^{KM}(d\bar{y}, d\bar{x}) \\ &= \left[ \sum_{i=1}^{n_1} W_{1,i:n_1} \hat{\xi}_1(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t, x) \right. \\ &\quad \left. - \sum_{l=1}^{n_0} W_{0,l:n_0} \hat{\xi}_0(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x) \right] \quad (10) \\ &= \frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{\xi}_1(Y_i, X_i, D_i; t, x) \delta_i}{1 - \hat{G}_1^{KM}(Q_i^-)} - \frac{\hat{\xi}_0(Y_i, X_i, D_i; t, x) \delta_i}{1 - \hat{G}_0^{KM}(Q_i^-)} \right), \end{aligned}$$

where  $\hat{G}_j^{KM}(\cdot)$  is the Kaplan and Meier (1958) estimator of  $G_j(\cdot)$ ,  $j = 0, 1$ , and the last equality follows from the results of Satten and Datta (2001).

From the above representation of  $\hat{I}(\cdot, \cdot)$ , one can clearly see that indeed the task of estimating  $I(\cdot, \cdot)$  is reduced to estimate  $p(\cdot)$ , use KM estimators for  $G_1(\cdot)$ ,  $G_0(\cdot)$ , and then applying plug-in principle. Moreover, in the absence of censoring, for  $i = 1, \dots, n$ ,  $Q_i = Y_i$ ,  $\delta_i = 1$  and  $W_{1,i:n_1} = W_{0,i:n_0} = n^{-1}$  a.s.. Therefore, (10) naturally reduces to (7). Hence, one can clearly see that our procedure is suitable regardless of the presence of censoring.

With  $\hat{I}(\cdot, \cdot)$  at hand, we are able to test the null hypothesis (1). Our test statistics are based on distances from  $\sqrt{n}\hat{I}(\cdot, \cdot)$  to zero. We consider the usual *sup* and  $L_2$  norms, leading

to the Kolmogorov-Smirnov (KS), and Cramér-von Mises (CvM) test statistics

$$KS_n = \sqrt{n} \sup_{(t,x) \in \mathcal{W}} \left| \hat{I}(t,x) \right|, \quad (11)$$

$$CvM_n = n \int_{\mathcal{W}} \left| \hat{I}(t,x) \right|^2 \hat{H}(dt,dx), \quad (12)$$

respectively, where  $\hat{H}(t,x)$  denotes the sample analog of  $H(t,x) = \mathbb{P}(Q \leq t, X \leq x)$ . Obviously, different test statistics could be developed by applying other distances, but for ease of exposition, we concentrate of  $KS_n$  and  $CvM_n$ .

Notice that, as a by-product of the our testing procedure, for  $t \in (-\infty, \tau_C)$ , one can estimate the unconditional distributional treatment effects (DTE)

$$\Upsilon(t) = \mathbb{E} \left[ \frac{D \delta 1 \{Q \leq t\}}{(1 - G_1(Q-)) p(X)} - \frac{(1 - D) \delta 1 \{Q \leq t\}}{(1 - p(X)) (1 - G_0(Q-))} \right] \quad (13)$$

by

$$\hat{\Upsilon}(t) = \frac{1}{n} \sum_{i=1}^n \left( \frac{D_i \delta_i 1 \{Q_i < \tau\}}{1 - \hat{G}_1^{KM}(Q_i-) \hat{p}(X_i)} - \frac{(1 - D_i) \delta_i 1 \{Q_i < \tau\}}{(1 - \hat{G}_0^{KM}(Q_i-)) (1 - \hat{p}(X_i))} \right).$$

Hubbard et al. (2000) proposes a similar estimator, but relying on parametric methods, whereas Abbring and van den Berg (2005) consider a related estimator in a context without covariates. A detailed comparison between these estimators is beyond the scope of this paper. Furthermore, by using test statistics similar to (11) and (12), one can test for the presence of overall treatment effects. To avoid repetition of arguments, we focus on the conditional tests.

## 3 Asymptotic Theory

### 3.1 Asymptotic linear representation

We now discuss the asymptotic theory for our test statistics  $KS_n$  and  $CvM_n$ , using the following notation. For a generic set  $\mathcal{G}$ , let  $l^\infty(\mathcal{G})$  be the Banach space of all uniformly bounded real functions on  $\mathcal{G}$  equipped with the uniform metric  $\|f\|_{\mathcal{G}} \equiv \sup_{z \in \mathcal{G}} |f(z)|$ . We

study the weak convergence of  $\sqrt{n} \left( \hat{I} - I \right) (\cdot, \cdot)$  and related processes as elements of  $l^\infty(\mathcal{W})$ . Let  $\Rightarrow$  denote weak convergence on  $(l^\infty(\mathcal{W}), \mathcal{B}_\infty)$  in the sense of J. Hoffmann-Jørgensen, where  $\mathcal{B}_\infty$  denotes the corresponding Borel  $\sigma$ -algebra - see e.g. van der Vaart and Wellner (1996).

As shown in Section 2.3,  $\hat{I}(\cdot, \cdot)$  is the difference of two empirical Kaplan-Meier integrals. However, because our KM integrals depend on a nonparametric estimate for the propensity score  $p(\cdot)$ , the results available in the literature cannot be straightforwardly applied, see e.g. Stute and Wang (1993b), Stute (1993, 1995, 1996), and Sellero et al. (2005). To accommodate this issue, we must present new results for our Kaplan-Meier integrals indexed by unknown, infinite-dimensional nuisance parameters. In short, we show that, due the propensity score estimation effect, an additional term in the asymptotic representation of  $\sqrt{n} \left( \hat{I} - I \right) (t, x)$  must be considered.

In order to proceed with the asymptotic analysis, let us introduce some additional notation. For  $j \in \{0, 1\}$ , let  $F_j(t|\bar{x}) = \mathbb{E}[1\{Y_j \leq t\} | X = \bar{x}]$ ,  $H_j(t) = \mathbb{P}(Q \leq t, D = j)$ ,  $H_{j,0}(y) = \mathbb{P}(Q \leq t, \delta = 0, D = j)$ , and  $H_{j,11}(t, x) = \mathbb{P}(Q \leq t, X \leq x, D = j, \delta = 1)$ . Note that  $H_j$ ,  $H_{j,0}$  and  $H_{j,11}$  may be consistently estimated from the observed data.

For  $j \in \{0, 1\}$  define

$$\gamma_{j,0}(\bar{t}) = \exp \left\{ \int_0^{\bar{t}-} \frac{H_{j,0}(d\bar{w})}{1 - H_j(\bar{w})} \right\}.$$

Let

$$\gamma_{j,1}(\bar{t}) = \frac{1}{1 - H_j(\bar{t})} \int 1\{\bar{t} < \bar{w}\} \xi_j(\bar{w}, \bar{x}, \bar{z}; t, x) \gamma_{j,0}(\bar{w}) H_{j,11}(d\bar{w}, d\bar{x})$$

and

$$\gamma_{j,2}(\bar{t}) = \int \int \frac{1\{\bar{v} < \bar{t}, \bar{v} < \bar{w}\} \xi_j(\bar{w}, \bar{x}, \bar{z}; t, x)}{[1 - H_j(\bar{v})]^2} \gamma_{j,0}(\bar{w}) H_{j,0}(d\bar{v}) H_{j,11}(d\bar{w}, d\bar{x}),$$

here  $\xi_1(\cdot, \cdot, \cdot; t, x)$  and  $\xi_0(\cdot, \cdot, \cdot; t, x)$  are as defined in (5) and (6), respectively. Put

$$\eta_{j,i}(t, x) = \xi_j(Q_i, X_i, D_i; t, x) \gamma_{j,0}(Q_i) \delta_i + \gamma_{j,1}(Q_i) (1 - \delta_i) - \gamma_{j,2}(Q_i). \quad (14)$$

Some remarks are necessary. First, the above representation relies only on the “known”

functions  $\xi_j$ ,  $j = 0, 1$ . Then, as discussed in Stute (1995, 1996), the first term of  $\eta_{j,i}(t, x)$  has expectation  $\mathbb{E}[\xi_j(Q, X, D; t, x)]$ . The second and third terms represent the estimation effect coming from not knowing  $G_j(\cdot)$  in (10), and they have identical expectations. Finally, notice that in the absence of censoring,  $\gamma_{j,0}(\cdot) = 1$  a.s., and  $\gamma_{j,1}(\cdot) = \gamma_{j,2}(\cdot) = 0$  a.s..

Given that  $\hat{I}(\cdot, \cdot)$  is the difference of empirical KM integrals, define

$$\eta_i(t, x) = \eta_{1,i}(t, x) - \eta_{0,i}(t, x), \quad (15)$$

the difference of (14) between the treated and control group.

To discuss the estimation effect coming from not knowing  $p(\cdot)$  in the KM-integrals, let

$$\alpha_1(X; t, x) = -p(X) 1\{X \leq x\} F_1(t|X), \quad (16)$$

$$\alpha_0(X; t, x) = (1 - p(X)) 1\{X \leq x\} F_0(t|X). \quad (17)$$

Notice that  $\alpha_1(\cdot; t, x)$  and  $\alpha_0(\cdot; t, x)$  are nothing more than the conditional expectation of the derivative of  $\xi_1$  and  $\xi_0$ , as defined in (5) and (6), with respect to  $p(\cdot)$ , respectively. Similarly to (15), define

$$\alpha(X; t, x) = \alpha_1(X; t, x) - \alpha_0(X; t, x). \quad (18)$$

Before presenting our asymptotic results, we need to assume some additional regularity conditions.

**Assumption 6** (i) *The support  $\chi_X$  of the  $k$ -dimensional covariate  $X$  is a Cartesian product of compact intervals,  $\chi_X = \prod_{j=1}^k [x_{lj}, x_{uj}]$ ;*

(ii) *The density of  $X$  is bounded, and bounded away from 0, on  $\chi_X$*

(iii) *For  $j \in \{0, 1\}$ , and any given  $t \in \chi_Y$ ,  $F_j(t|x)$  is continuously differentiable in  $x \in \chi_X$ .*

**Assumption 7** *For all  $x \in \chi_X$ , the propensity score  $p(x)$  is continuously differentiable of order  $s \geq 13k$ , where  $k$  is the dimension of  $X$ .*

**Assumption 8** *The series logit estimator of  $p(x)$  uses a power series with  $L = a \cdot N^v$  for some  $a > 0$  and  $1/(s/k - 2) < v < 1/11$ .*

Similar assumptions have been done by Hahn (1998), Hirano et al. (2003), Crump et al. (2008), Donald and Hsu (2013), among others. Assumption 6 restrict the distribution of  $X$  and  $Y$  and requires that all covariates are continuous. By imposing these restrictions, we are able to use Newey (1997) results for series estimators. Nonetheless, at the expense of additional notation, we can deal with the case where  $X$  has both continuous and discrete components by means of sample splitting based on the discrete covariates. In order to avoid cumbersome notation, we abstract from this point in the rest of the paper. Assumption 7 requires sufficient smoothness of the propensity score, whereas Assumption 8 restrict the rate at which additional terms are added to the series approximation of  $p(x)$ , depending on the dimension of  $X$  and the number of derivatives of  $p(x)$ . The restriction on the derivatives in Assumption 7 guarantees the existence of a  $v$  that satisfy the conditions in Assumption 8.

Under the aforementioned conditions, we can state our first asymptotic result, which provides the representation of  $\sqrt{n}(\hat{I} - I)(t, x)$  over  $\mathcal{W}$ .

**Lemma 1** *Under Assumptions 1-8, we have*

$$\sup_{(t,x) \in \mathcal{W}} \left| \sqrt{n} \left( \hat{I}(t, x) - I(t, x) \right) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ [\eta_i(t, x) - I(t, x)] + \alpha(X_i; t, x) (D_i - p(X_i)) \} \right| = o(1).$$

Lemma 1 shows that the estimator  $\hat{I}(t, x)$  can be represented as asymptotically linear:

$$\hat{I}(t, x) = I(t, x) + \frac{1}{n} \sum_{i=1}^n \{ \psi_i(t, x) + \tilde{\alpha}_i(t, x) \} + o_{\mathbb{P}}(n^{-1/2})$$

where

$$\psi_i(t, x) = \eta_i(t, x) - I(t, x),$$

$\eta_i(t, x)$  being defined as in (15) and

$$\tilde{\alpha}_i(t, x) = \alpha(X_i; t, x)(D_i - p(X_i)). \quad (19)$$

The known-IPW estimator, (10) with  $\hat{p}(x)$  replaced by  $p(x)$ , is asymptotically linear with score function  $\psi(\cdot, \cdot)$ . The function  $\tilde{\alpha}(t, x)$  represents the effect on the score function of estimating  $p(\cdot)$ .

### 3.2 Asymptotic null distribution

Using the uniform representation of Lemma 1, we next derive the weak convergence of the processes  $\sqrt{n}\hat{I}(t, x)$  under the null hypothesis (1).

**Theorem 1** *Under the null hypothesis (1) and Assumptions 1-8, we have*

$$\sqrt{n}\hat{I}(t, x) \Rightarrow C_\infty,$$

where  $C_\infty$  is Gaussian process with zero mean and covariance function

$$V((t_1, x_1), (t_2, x_2)) = \mathbb{E}[\{\psi(t_1, x_1) + \tilde{\alpha}(t_1, x_1)\} \{\psi(t_2, x_2) + \tilde{\alpha}(t_2, x_2)\}]. \quad (20)$$

Now, we can apply the continuous mapping theorem in order to characterize the limiting null distributions of our test statistics using the sup and  $L_2$  distances.

**Corollary 1** *Under the null hypothesis (1) and the assumptions of Theorem 1,*

$$\begin{aligned} KS_n &\xrightarrow{d} \sup_{(t,x) \in \mathcal{W}} |C_\infty(t, x)|, \\ CvM_n &\xrightarrow{d} \int_{\mathcal{W}} |C_\infty(t, x)|^2 H(dt, dx). \end{aligned}$$

Let  $T_n$  be a generic notation for  $KS_n$  and  $CvM_n$ . From Corollary 1, it follows immediately that

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{T_n > c_\alpha^T\right\} = \alpha$$

where

$$c_\alpha^T = \inf \left\{ c \in [0, \infty) : \lim_{n \rightarrow \infty} \mathbb{P} \{ T_n > c \} = \alpha \right\},$$

### 3.3 Asymptotic distribution under fixed and local alternatives

Now, we analyze the asymptotic properties of our tests under the fixed alternative  $H_1$ . Under  $H_1$ , there is at least one  $(t, x) \in \mathcal{W}$  such that  $\Upsilon(t, x) \neq 0$ , implying that  $I(t, x) \neq 0$  for some  $(t, x) \in \mathcal{W}$ . Therefore, our test statistics  $KS_n$  and  $CvM_n$  diverge to infinity. Given that the critical values are bounded, it follows that our tests are consistent. We formalize this result in the next theorem.

**Theorem 2** *Under Assumptions 1-8 and the alternative hypothesis  $H_1$ , we have*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ KS_n > c_\alpha^{KS} \right\} &= 1, \\ \lim_{n \rightarrow \infty} \mathbb{P} \left\{ CvM_n > c_\alpha^{CvM} \right\} &= 1. \end{aligned}$$

Given that our test statistics diverge to infinity under fixed alternatives, it is desirable studying the asymptotic power of these tests under local alternatives. To this end, we study the asymptotic behavior of  $\hat{I}(t, x)$  under alternative hypotheses converging to the null at the parametric rate  $n^{-1/2}$ .

Consider the following class of local alternatives:

$$H_{1,n} : \Upsilon(t, x) = \frac{1}{\sqrt{n}} h(t, x) \quad \forall (t, x) \in \mathcal{W}, \quad (21)$$

In the sequel, we need that (21) satisfies the following regularity condition.

**Assumption 9** (a)  $h(\cdot, \cdot)$  is an  $F$ -integrable function;

(b) the set  $h_n \equiv [(t, x) \in \mathcal{W} : h(t, x) \neq 0]$  has positive Lebesgue measure.

**Theorem 3** *Under the local alternatives (21) and Assumptions 1- 9,*

$$\sqrt{n} \hat{I}(t, x) \Rightarrow C_\infty + R$$

where  $C_\infty$  is the process defined in Theorem 1 and  $R(\cdot)$  is the deterministic function

$$R(t, x) = \mathbb{E}[h(t, X) (p(X) (1 - p(X))) 1_{\{X \leq x\}}].$$

The following corollary is a consequence of the continuous mapping theorem and Theorem 3.

**Corollary 2** *Under the local alternatives (21), and Assumptions 1- 9,*

$$\begin{aligned} KS_n &\xrightarrow{d} \sup_{(t,x) \in \mathcal{W}} |C_\infty(t, x) + R(t, x)|, \\ CvM_n &\xrightarrow{d} \int_{\mathcal{W}} |C_\infty(t, x) + R(t, x)|^2 H(dt, dx), \end{aligned}$$

From the above corollary, we see that our test statistics, under local alternatives of the form of (21), converge to a different distribution due to the presence of a deterministic shift function  $R$ . This additional term guarantees the good local power property of our test.

### 3.4 Estimation of critical values

From the above theorems, we see that the asymptotic distribution of  $\sqrt{n}\hat{I}(\cdot, \cdot)$  depends on the underlying data generating process and standardization is complicated in this case. Therefore, we propose a bootstrap method to estimate the critical values of our test. Our bootstrap procedure is related to the wild bootstrap, but instead of just resampling imposing the restriction under  $H_0$ , we use the asymptotic linear representation of  $\sqrt{n}\hat{I}(\cdot, \cdot)$ . More precisely, we consider the multiplier-type bootstrap as Stute et al. (2000), Delgado and González-Manteiga (2001), Barrett and Donald (2003) and Donald and Hsu (2013) suggest in different contexts. The proposed procedure has good theoretical and empirical properties, is straightforward to verify its asymptotic validity, and is computationally easy to implement.

In order to implement the bootstrap, we need nonparametric estimators for the terms in the asymptotic linear representation of Lemma 1, namely the propensity score  $p(\cdot)$ ,  $\eta(t, x)$  as defined in (15), and  $\alpha(\cdot; t, x)$  as in (18).

As already discussed, we estimate  $p(\cdot)$  using the SLE of Hirano et al. (2003). In order to

estimate  $\eta(t, x)$ , we notice that after plugging in  $\hat{p}(\cdot)$ , each  $\gamma$  only depends on  $H$ -functions and is therefore estimable just replacing the  $H$ -terms by their empirical counterparts. Then, we estimate  $\eta(t, x)$  by its empirical analogue,

$$\hat{\eta}(t, x) = \hat{\eta}_1(t, x) - \hat{\eta}_0(t, x)$$

such that, for  $j = 0, 1$ ,

$$\begin{aligned}\hat{\eta}_j(t, x) &= \hat{\xi}_j(Q, X, D; t, x) \hat{\gamma}_{j,0}(Q) \delta_{j,i} + \hat{\gamma}_{j,1}(Q) (1 - \delta) - \hat{\gamma}_{j,2}(Q), \\ \hat{\gamma}_{j,0}(\bar{t}) &= \exp \left\{ \int_0^{\bar{t}-} \frac{\hat{H}_{j,0}(d\bar{w})}{1 - \hat{H}_j(\bar{w})} \right\}, \\ \hat{\gamma}_{j,1}(\bar{t}) &= \frac{1}{1 - \hat{H}_j(\bar{t})} \int 1 \{ \bar{t} < \bar{w} \} \hat{\xi}_j(\bar{w}, \bar{x}, \bar{z}; t, x) \hat{\gamma}_{j,0}(\bar{w}) \hat{H}_{j,11}(d\bar{w}, d\bar{x}), \\ \hat{\gamma}_{j,2}(\bar{t}) &= \int \int \frac{1 \{ \bar{v} < \bar{t}, \bar{v} < \bar{w} \} \hat{\xi}_j(\bar{w}, \bar{x}, \bar{z}; t, x)}{[1 - \hat{H}_j(\bar{v})]^2} \hat{\gamma}_{j,0}(\bar{w}) \hat{H}_{j,0}(d\bar{v}) \hat{H}_{j,11}(d\bar{w}, d\bar{x}),\end{aligned}$$

where  $\hat{\xi}_1(\cdot, \cdot, \cdot; t, x)$  and  $\hat{\xi}_0(\cdot, \cdot, \cdot; t, x)$  are defined in (8) and (9), respectively, and

$$\begin{aligned}\hat{H}_j(\bar{w}) &= \frac{1}{n} \sum_{i=1}^n 1 \{ Q_i \leq \bar{w} \} 1 \{ D_i = j \}, \\ \hat{H}_{j,0}(\bar{w}) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) 1 \{ Q_i \leq \bar{w} \} 1 \{ D_i = j \} \\ \hat{H}_{j,11}(\bar{w}, \bar{x}) &= \frac{1}{n} \sum_{i=1}^n \delta_i 1 \{ Q_i \leq \bar{w} \} 1 \{ X_i \leq \bar{x} \} 1 \{ D_i = j \}\end{aligned}$$

are the empirical counterparts of  $H_j(\bar{w})$ ,  $H_{j,0}(\bar{w})$  and  $H_{j,11}(\bar{w})$ , respectively.

Finally, we must consider nonparametric estimate for  $\alpha(X; t, x) = \alpha_1(X; t, x) - \alpha_0(X; t, x)$ ,  $\alpha_1(X; t, x)$  and  $\alpha_0(X; t, x)$  being defined in (16) and (17), respectively. To this end, notice that

$$\begin{aligned}\alpha(X; t, x) &= -\mathbb{E} [D 1 \{ Y \leq t \} + (1 - D) 1 \{ Y \leq t \} | X] 1 \{ X \leq x \} \\ &= -\mathbb{E} [1 \{ Y \leq t \} | X] 1 \{ X \leq x \}.\end{aligned}$$

If we fully observe  $(Y, X, D)$ , we could estimate this conditional expectation using nonpara-

metric series regression of  $1\{Y \leq \cdot\}$  on  $X$ , as similarly adopted by Hirano et al. (2003) and Donald and Hsu (2013). Given that the outcome of interest  $Y$  is subjected to censoring, such procedure is not at our disposal. To the best of our knowledge, no nonparametric estimator for  $\alpha(\cdot; t, x)$  is yet available.

Notwithstanding, by using the Kaplan-Meier weights as discussed in Sections 2 and 3.1, we can overcome such problem and estimate  $\alpha(\cdot; t, x)$  by the Kaplan-Meier series estimator

$$\begin{aligned} & \hat{\alpha}^{KM}(X; t, x) \\ &= - \left( \frac{1}{n} \sum_{i=1}^n \left( \frac{D_i}{1 - \hat{G}_1^{KM}(Q_{i-})} + \frac{1 - D_i}{1 - \hat{G}_0^{KM}(Q_{i-})} \right) \delta 1\{Q_i \leq t\} R^L(X_i) \right)' \\ & \quad \times \left( \frac{1}{n} \sum_{i=1}^n R^L(X_i) R^L(X_i)' \right)^{-1} R^L(X) 1\{X \leq x\}, \end{aligned}$$

where  $R^L(\cdot)$  is the same power series used in SLE estimator, with potentially different number of series. The uniform consistency of the aforementioned nonparametric estimator for  $\alpha(X; t, x)$  is proved in Lemma A.4 in Appendix A.

Once we have nonparametric estimators  $p(\cdot)$ ,  $\eta(t, x)$ , and  $\alpha(\cdot; t, x)$ , the bootstrapped version of  $\hat{I}(t, x)$  is given by

$$\hat{I}^*(t, x) = \frac{1}{n} \sum_{i=1}^n [\hat{\eta}_i(t, x) + \hat{\alpha}^{KM}(X_i; t, x) (D_i - \hat{p}(X_i))] V_i$$

where the random variables  $\{V_i\}_{i=1}^n$  are *iid* as a random variable  $V$  with bounded support, zero mean and variance one, being independent generated from the sample  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^N$ .

Replacing  $\hat{I}(t, x)$  with  $\hat{I}^*(t, x)$ , we get the bootstrap versions of  $KS_n$  and  $CvM_n$ ,  $KS_n^*$  and  $CvM_n^*$ , respectively. The asymptotic critical values are estimated by

$$\begin{aligned} c_{n,\alpha}^{KS,*} &\equiv \inf \left\{ c_\alpha \in [0, \infty) : \lim_{n \rightarrow \infty} \mathbb{P}_n^* \{KS_n^* > c_\alpha\} = \alpha \right\}, \\ c_{n,\alpha}^{CvM,*} &\equiv \inf \left\{ c_\alpha \in [0, \infty) : \lim_{n \rightarrow \infty} \mathbb{P}_n^* \{CvM_n^* > c_\alpha\} = \alpha \right\} \end{aligned}$$

where  $\mathbb{P}_n^*$  means bootstrap probability, i.e. conditional on the sample  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^n$ . In practice,  $c_{n,\alpha}^{KS,*}$  and  $c_{n,\alpha}^{CvM,*}$  are approximated as accurately as desired by  $(KS_n^*)_{B(1-\alpha)}$

and  $(CvM_n^*)_{B(1-\alpha)}$ , the  $B(1-\alpha)$ -th order statistic from  $B$  replicates  $\{KS_n^*\}_{l=1}^B$  of  $KS_n^*$  or  $\{CvM_n^*\}_{l=1}^B$  of  $CvM_n^*$ , respectively.

The next theorem proves the validity of the proposed multiplicative bootstrap. Notice that we need an additional smoothness assumption on  $F_j(\cdot|X)$ ,  $j \in \{0, 1\}$ .

**Theorem 4** *Let Assumptions 1-9 hold. Additionally, for  $j \in \{0, 1\}$ , assume that  $F_j(\cdot|X)$  is continuously differentiable of order  $m \geq k$ , where  $k$  is the dimension of  $X$ . Assume  $\{V_i\}_{i=1}^n$  are iid, independent of the sample  $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^N$ , bounded with zero mean and variance one. Then, under the null hypothesis (1), any fixed alternative hypothesis or under the local alternatives (21)*

$$\sqrt{n}\hat{I}_*^* \Rightarrow C_\infty$$

where  $C_\infty$  is the same Gaussian process of Theorem 1 and  $\Rightarrow$  denoting weak convergence a.s. under the bootstrap law (see Giné and Zinn (1990)).

Straightforward application of the continuous mapping theorem lead us to conclude that our bootstrap-based tests has correct asymptotic size, are consistent against fixed alternatives and are able to detect contiguous alternatives of the form of (21).

## 4 Monte Carlo simulations

In this section, we conduct a small scale Monte Carlo exercise in order to study the finite sample properties of our test statistics for the null hypothesis (1). The  $\{V_i\}_{i=1}^n$  used in the bootstrap implementations are independently generated as  $V$  with  $\mathbb{P}(V = 1 - \kappa) = \kappa/\sqrt{5}$  and  $\mathbb{P}(V = \kappa) = 1 - \kappa/\sqrt{5}$ , where  $\kappa = (\sqrt{5} + 1)/2$ , as proposed by Mammen (1993). The bootstrap critical values are approximated by Monte Carlo using 1000 replications and the simulations are based on 10000 Monte Carlo experiments. We report rejection probabilities at the 5% significance level. Results for 10% and 1% significance levels are similar and available upon request.

We consider the following four designs:

$$(i). Y_0 \sim \text{Exponential} \left( \frac{1}{1.1 + X} \right), Y_1 \sim \text{Exponential} \left( \frac{1}{1.1 + X} \right),$$

$$C_1 = C_2 \sim a_{11} + b_{11} \times \text{Exponential}(1);$$

$$(ii). Y_0 \sim \text{Exponential} \left( \frac{1}{1.1 + X} \right), Y_1 \sim \text{Exponential} \left( \frac{1}{1.1 + X} \right),$$

$$C_1 \sim a_{12} + b_{12} \times \text{Exponential}(1); C_2 \sim a_{22} + b_{22} \times \text{Exponential}(1);$$

$$(iii). Y_0 \sim \text{Exponential} \left( \frac{1}{1.1 + X} \right), Y_1 \sim \text{Exponential} \left( \frac{1}{0.1 + X} \right),$$

$$C_1 \sim a_{13} + b_{13} \times \text{Exponential}(1), C_2 \sim a_{23} + b_{23} \times \text{Exponential}(1);$$

$$(iv). Y_0 \sim \text{Exponential} \left( \frac{1}{1.1 + 2X} \right), Y_1 \sim \text{Exponential} \left( \frac{1}{0.1 + x} \right),$$

$$C_1 \sim a_{14} + b_{14} \times \text{Exponential}(1), C_2 \sim a_{24} + b_{24} \times \text{Exponential}(1);$$

where  $X$  is distributed as  $U[0, 1]$ , independently of  $Y_0, Y_1, C_1$  and  $C_0$ , and the parameters  $a$  and  $b$  are chosen such that the percentage of censoring is equal to 0, 10 or 30 percent in the whole sample. Design (i) and (ii) fall under the null hypothesis, and designs (iii) – (iv) fall under the alternative. Design (i) differs from design (ii) by the censoring distribution: in (i), the censoring variable is the same for treated and control group, whereas in design (ii)  $C_1$  and  $C_2$  follow different distributions. In design (ii) we set that the censoring level under treated and control groups are different: it is 0, 5, and 20 under control and 0, 15 and 40 under treatment. For the other designs, the censoring proportion is equal for the treatment and control groups. In design (iii), the CDTE does not depend on covariates, whereas in design (iv) it does. In all designs,  $\mathbb{P}(D = 1|X) = X$ .

We report the proportion of rejections for sample sizes  $n = 100, 300$  and  $1000$ . We estimate  $p(\cdot)$  using the SLE: with  $n = 100$  we use  $1, X, X^2$ , with  $n = 300$  we use  $1, X, X^2, X^3$  and with  $n = 1000$  we use  $1, X, X^2, X^3, X^4, X^5$  as power functions in the estimation proce-

dure.

We compare our proposed tests  $KS_n$  and  $CvM_n$  as in (11)-(12), with two others alternatives: the ‘naive’ procedure where censoring is ignored ( $KS_n^{naive}$  and  $CvM_n^{naive}$ ), and the analogous procedure of  $KS_n$  and  $CvM_n$  but imposing that the censoring variable is the same under treatment and control groups ( $KS_n^{same}$  and  $CvM_n^{same}$ ), both implemented with the assistance of a bootstrap analogous to the one discussed in Section 3.4. The proportion of rejections are presented in Table 1.

Table 1: Empirical Rejection probabilities.

Design	Tests	n=100			n=300			n=1000		
		% of Censoring			% of Censoring			% of Censoring		
		0%	10%	30%	0%	10%	30%	0%	10%	30%
(i)	$KS_n$	-	5.02	4.78	-	5.13	4.87	-	5.22	5.00
	$CvM_n$	-	5.47	5.22	-	5.39	5.01	-	5.22	5.31
	$KS_n^{naive}$	5.02	5.34	3.34	4.96	4.98	3.09	5.55	6.19	5.53
	$CvM_n^{naive}$	4.96	5.34	4.17	5.17	4.88	3.74	5.47	6.15	5.29
	$KS_n^{same}$	-	4.85	4.32	-	5.13	5.28	-	5.14	5.38
	$CvM_n^{same}$	-	5.32	4.89	-	5.44	5.09	-	5.26	5.22
(ii)	$KS_n$	-	5.01	5.14	-	5.08	5.17	-	5.33	5.02
	$CvM_n$	-	5.39	5.44	-	5.28	4.95	-	5.30	5.12
	$KS_n^{naive}$	4.98	5.65	8.79	5.53	6.19	24.85	5.23	12.40	87.00
	$CvM_n^{naive}$	4.88	5.74	6.33	5.29	6.15	12.06	5.19	8.72	43.33
	$KS_n^{same}$	-	6.39	19.78	-	11.21	60.18	-	43.79	98.84
	$CvM_n^{same}$	-	5.81	12.26	-	6.82	33.32	-	12.03	89.46
(iii)	$KS_n$	-	83.97	60.99	-	99.83	92.35	-	100	99.27
	$CvM_n$	-	93.00	87.23	-	100	99.98	-	100	100
	$KS_n^{naive}$	88.14	84.91	68.28	99.98	99.97	99.32	100	100	100
	$CvM_n^{naive}$	93.98	90.47	78.34	100	100	99.80	100	100	100
	$KS_n^{same}$	-	78.31	38.22	-	99.52	86.27	-	100	98.69
	$CvM_n^{same}$	-	90.17	77.64	-	100	99.90	-	100	100
(iv)	$KS_n$	-	93.49	73.51	-	99.92	94.75	-	100	99.64
	$CvM_n$	-	98.34	95.13	-	100	100	-	100	100
	$KS_n^{naive}$	96.92	94.88	82.97	100	100	99.94	100	100	100
	$CvM_n^{naive}$	98.74	97.22	89.36	100	100	99.99	100	100	100
	$KS_n^{same}$	-	77.64	49.75	-	99.81	90.71	-	100	100
	$CvM_n^{same}$	-	97.25	89.60	-	100	100	-	100	100

Note: One thousand bootstrap replications. Ten thousand Monte Carlo simulations. 5% level.

We observe that our tests  $KS_n$  and  $CvM_n$  exhibits good size accuracy for both designs (i) and (ii) even when  $n = 100$ . In design (i), tests based on the ‘naive’ and the ‘common

censoring' control size, though once we increase the sample size and the censoring proportion, the size of  $KS_n^{naive}$  and  $CvM_n^{naive}$  fall below the nominal level. Although one may find the result that the 'naive' procedure is able to control the Type-I error surprising, the reason behind this is simple: since  $Y_1 = Y_0$ , and at the same time  $C_1 = C_0$ , the censored outcomes  $Q_1 = \min(Y_1, C_1)$  and  $Q_0 = \min(Y_0, C_0)$  are also equal. Nonetheless, when the  $C_1$  is different than  $C_2$ , as in design (ii), this is not true anymore. As one can see from Table 1, the tests procedures that either ignore the censoring or incorrectly impose the assumption of common  $G$ 's are not able to control size in this situation. This size distortions become more evident as we increase the sample size and the censoring level, reaching values higher than 80%.

With respect to power, our tests  $KS_n$  and  $CvM_n$  reach moderate levels for  $n = 100$ , but they uniformly increase and reach satisfactory levels when sample size is 300. The power is decreasing with the degree of censoring. For the considered designs,  $CvM_n$  tends to have higher power than  $KS_n$ . In addition, we can see that our proposed tests has similar and some times even higher power to those based on the 'naive' and the 'common censoring' procedures. Overall, these simulations show that the proposed bootstrap tests  $KS_n$  and  $CvM_n$  exhibit very good size accuracy and power for relatively small sample sizes. On other hand, the tests  $KS_n^{naive}$ ,  $CvM_n^{naive}$ ,  $KS_n^{same}$  and  $CvM_n^{same}$  may not be reliable due to their inability of controlling size when the censoring distributions differ in the two treatment regimes.

## 5 Some applications of the basic setup

### 5.1 Average treatment effects

So far, we have only discussed tests for the existence of distributional treatment effects. Although the proposed tests for zero distributional treatment effects are able to detect a very broad set of alternative hypotheses, we are still not able to pin down the direction of the departure from the null hypothesis of interest. For instance, if we reject the null of zero distributional treatment effect for all subpopulations defined by covariates, we unfortunately

do not know if the policy has affect the conditional mean or, instead, any other particular feature of the outcome distribution. Given that the policy evaluation literature has given a great deal of importance to the average treatment effect, in this section we show how one adapt our tests to focus on this particular measure.

Let  $\Upsilon^{CATE}(x) = \mathbb{E}[Y_1 - Y_0|X = x]$ . Remember that, as discussed in Section 2, we may be unable to test hypotheses concerning  $\Upsilon^{CATE}(x)$  itself because of lack of information in the right tail of the outcome distribution due to the censoring mechanism. Therefore, we focus our attention to the trimmed versions of  $\Upsilon^{CATE}(x)$ ,  $\Upsilon_\tau^{CATE}(x) = \mathbb{E}[Y_1 1\{Y_1 < \tau\} - Y_0 1\{Y_0 < \tau\} | X = x]$ , where  $\tau \leq \tau_C$ . Similar procedures have been previously considered by Sello et al. (2005) and Pardo-Fernandez and Van Keilegom (2006).

We are concerned with the following hypothesis:

$$H_0^{CATE} : \Upsilon_\tau^{CATE}(x) = 0 \quad \forall x \in \mathcal{W}_X \quad (22)$$

where  $\mathcal{W}_X \subseteq \chi_x$ ,  $\chi_x$  denoting the support of  $X$ . Under  $H_0^{CATE}$ , the trimmed average treatment effect (ATE) is equal to zero for all subpopulations defined by covariates. The alternative hypothesis is the negation of the null.

Following the same steps as in Section 2, our Kolmogorov-Smirnov (KS) type test statistic for hypothesis (22) is

$$KS_n^{CATE} = \sup_{x \in \mathcal{W}_X} \left| \sqrt{n} \hat{I}_\tau^{CATE}(x) \right|, \quad (23)$$

where  $\hat{I}_\tau^{CATE}(x)$  is defined as

$$\hat{I}_\tau^{CATE}(x) = \frac{1}{n} \sum_{i=1}^n \left( \frac{D_i(1 - \hat{p}(X_i))}{1 - \hat{G}_1^{KM}(Q_{i-})} - \frac{(1 - D_i)\hat{p}(X_i)}{1 - \hat{G}_0^{KM}(Q_{i-})} \right) \delta_i Q_i 1\{Q_i < \tau\} 1\{X_i \leq x\}.$$

The discussion for the Cramér-von Mises test is the same and is therefore omitted. Notice that when  $\tau = \tau_C$ ,  $\delta 1\{Q < \tau\} = \delta$ , and therefore no user-chosen trimming is necessary. This is of particular importance because, in this case, we are using all the information about the average treatment effect available in the data.

In order to proceed with the asymptotic analysis, we need the following integrability

assumption, which guarantees that the variance of our proposed estimators is finite and that the censoring effects do not dominate in the right tails. See Stute (1996) for a detailed discussion.

**Assumption 10** For  $j \in \{0, 1\}$ , assume the following integrability condition

$$\begin{aligned} \mathbb{E} [(Q_j \gamma_{j,0}(Q))^2] &< \infty, \\ \mathbb{E} [ |Y_j| C_j^{1/2}(Y) ] &< \infty, \end{aligned} \tag{24}$$

where

$$C_j(w) = \int_{-\infty}^{w-} \frac{G_j(dy)}{[1 - H_j(y)][1 - G_j(y)]}$$

For a given  $\tau \leq \tau_C$ , consider the class of local alternatives

$$H_{1,n}^{CATE} : \Upsilon_{\tau}^{CATE}(x) = \frac{1}{\sqrt{n}} h^{CATE}(x) \quad \forall x \in \mathcal{W}_X, \tag{25}$$

that satisfy the following regularity condition.

**Assumption 11** Assume that

- (a)  $h^{CATE}(\cdot)$  is an  $F$ -integrable function;
- (b) the set  $h_n^{CATE} \equiv \left[ x \in \mathcal{W}_X : h^{CATE}(x) \neq 0 \right]$  has positive Lebesgue measure.

Using an analogous procedure described in Section 3.4, let  $c_{\alpha,n}^{CATE,*}$  denote the bootstrap critical value of the  $KS_n^{CATE}$ . In the next theorem, we prove that, for a given  $\tau$ , our tests for CATE are asymptotically unbiased, consistent and are able to detect local alternatives of the form of (25).

**Theorem 5** Suppose Assumptions 1-8, 10 and 11 hold. Additionally, assume that for  $j \in \{0, 1\}$ ,  $\mathbb{E}(Y_j|X)$  is continuously differentiable of order  $m \geq k$ , where  $k$  is the dimension of  $X$ . Then, for a fixed  $\tau \leq \tau_C$ ,

1. Under  $H_0^{CATE}$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha,n}^{CATE,*} \right\} = \alpha$ .

2. Under  $H_1^{CATE}$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha, n}^{CATE, *} \right\} = 1$ .
3. Under  $H_{1, n}^{CATE}$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha, n}^{CATE, *} \right\} > \alpha$ .

From the above discussion, we conclude that with a simple modification of our tests for distributional treatment effects, we can concentrate on tests for average treatment effects. In general, these tests can complement each other.

## 5.2 Testing within the Local Treatment Effect setup

The goal of this section is to show that, in case the unconfoundedness assumption does not hold, that is, if Assumption 2 fails, our tests are still applicable to the local average treatment effect (LTE) setup of Imbens and Angrist (1994) and Angrist et al. (1996).

The LTE setup presumes the availability of a binary instrumental variable  $Z$  for the treatment assignment. Denote  $D_0$  and  $D_1$  the value that  $D$  would have taken if  $Z$  is equal to zero or one, respectively. The realized treatment is  $D = ZD_1 + (1 - Z)D_0$ .

In order to identify the LTE for the subpopulation of compliers, that is, individuals who comply with their actual assignment of treatment and would have complied with the alternative assignment, we need the following assumption.

**Assumption 12** (i)  $(Y_0, Y_1, D_1, D_0, C_1, C_0) \perp\!\!\!\perp Z|X$ .

(ii)  $\forall x \in \mathcal{W}$ , and some  $\varepsilon > 0$ ,

$$\varepsilon < \mathbb{P}(Z|X = x) \equiv q(x) < 1 - \varepsilon,$$

and

$$\mathbb{P}(D_1 = 1|X = x) > \mathbb{P}(D_0 = 1|X = x) \forall x \in \mathcal{W}_X,$$

(iii)  $\mathbb{P}(D_1 > D_0|X = x) = 1 \forall x \in \mathcal{W}$ .

The null hypothesis of interest in this setup is

$$H_0^L : \Upsilon^L(t, x) = 0 \forall (t, x) \in \mathcal{W},$$

where

$$\Upsilon^L(t, x) = \mathbb{E}[1\{Y_1 \leq t\} - 1\{Y_0 \leq t\} | X = x, Pop = Comp].$$

This hypothesis is the analogous of (1) within the LTE setup.

In order to proceed, we must express  $\Upsilon^L(t, x)$  in terms of  $(Q, \delta, D, X)$ . It turns out that, under Assumptions 3-5 and 12, for  $(t, x) \in (-\infty, \tau_C) \times (-\infty, \infty)^k$ ,

$$\Upsilon^L(t, x) = \mathbb{E} \left[ \frac{Z1\{Y \leq t\}}{1 - G_1(Q-)q(X)} - \frac{(1 - Z)1\{Y \leq t\}}{(1 - q(X))1 - G_0(Q-)} | X = x \right] \\ \Bigg/ \mathbb{E} \left[ \frac{ZD}{q(X)} - \frac{(1 - Z)D}{1 - q(X)} | X = x \right]$$

From Assumption 12, the denominator of  $\Upsilon^L(\cdot, \cdot)$  is always strictly positive. Therefore, from the discussion in Section 2, the hypothesis of zero conditional distribution treatment effect among compliers can be equivalently written as

$$H_0^L : I^L(t, x) = 0 \forall (t, x) \in \mathcal{W},$$

where

$$I^L(t, x) = \mathbb{E} \left[ \left( \frac{Z(1 - q(X))}{1 - G_1(Q-)} - \frac{(1 - Z)q(X)}{1 - G_0(Q-)} \right) \delta 1\{Q \leq t\} 1\{X \leq x\} \right].$$

Noticing that once we replace  $Z$  to  $D$ , and  $q(x)$  to  $p(x)$ ,  $I^L(t, x)$  is equal  $I(t, x)$ , that is, the LTE framework reduces to the unconfounded framework. Therefore, we conclude that our tests for zero treatment effects with censored data are valid not only when the treatment assignment is unconfounded, but also to a particular case when the selection to treatment is based on unobservables, namely the local treatment effect setup of Imbens and Angrist (1994) and Angrist et al. (1996).

### 5.3 Dynamic treatment assignments

Until now, all proposed tests rely on individuals entering the treatment at the beginning of the spell. Nonetheless, this setup might be restrictive for some important applications

where the treatment might start at any time. A leading example is the active labor market policy (ALMP) programs for the unemployed. The common feature of ALMP is that participation is not instantaneous upon inflow into unemployment, but individuals are observed to enter ALMP programs at any possible time since the start of the unemployment spell. This dynamic selection mechanism introduces some potential problems to select a proper control group. The main issue within this dynamic setup is that individuals currently non-treated might become treated later. Given that the probability of enrollment increases with the elapsed duration, the treatment status depends on the outcome, and therefore, unconfoundedness-based tests like ours may not be suitable. Nonetheless, in this subsection we show that, by focusing on the effect of treatment now versus continuing to wait for treatment, as initially proposed by Sianesi (2004), our test statistics are still suitable.

In order to formalize this idea, we need to introduce some additional notation. The eligible population at time  $u$  are those still in the state of interest after  $u$  periods. For the eligibles at  $u$ , denote  $D^{(u)} = 1$  for joining a program at  $u$  and  $D^{(u)} = 0$  for not joining at least up to  $u$ . Denote  $Y_1^{(u)}$  and  $Y_0^{(u)}$  as the potential outcomes if treated at  $u$  and not yet treated up to  $u$ , respectively. Note that the potential outcomes  $Y_1^{(u)}$  and  $Y_0^{(u)}$  are only defined for those who are still in the state of the interest at time  $u$ , that is, only for those  $Y_1^{(u)} > u$ ,  $Y_0^{(u)} > u$ . Assume that  $\mathbb{P}(Y_1^{(u)} > u|X)$  and  $\mathbb{P}(Y_0^{(u)} > u|X)$  is always between  $\varepsilon$  and  $1 - \varepsilon$ , for some  $\varepsilon > 0$ .

The conditional distributional treatment effect is given by

$$\Upsilon^{(u)}(t, x) = \mathbb{E} \left[ 1\{Y_1^{(u)} \leq t\} - 1\{Y_0^{(u)} \leq t\} | X = x, Y_1^{(u)} > u, Y_0^{(u)} > u \right].$$

Under Assumptions 2-5, but with  $D^{(u)}$ ,  $Y_1^{(u)}$  and  $Y_0^{(u)}$  playing the role of  $D$ ,  $Y_1$  and  $Y_0$ , we have that

$$\Upsilon^{(u)}(t, x) = \mathbb{E} \left[ \left( \frac{D^{(u)} \delta 1\{Q \leq t\}}{(1 - G_1(Q-)) p(X)} - \frac{(1 - D^{(u)}) \delta 1\{Q \leq t\}}{(1 - p(X)) (1 - G_0(Q-))} \right) | X = x, Q > u \right], \quad (26)$$

for  $(t, x) \in (-\infty, \tau_C) \times (-\infty, \infty)^k$ .

Notice that (26) is nothing more than (2) restricted to the subpopulation of those who are still at the state of interest at time  $u$ . Therefore, once this restriction is applied to the observed data, all the test statistics previously described can be straightforwardly used. Hence, we conclude that our proposal is also suitable for the case of dynamic treatment assignment.

## 6 Evaluation of labor market programs

In this section, we demonstrate that our proposed tests can be useful in practice. We consider one application with experimental data, the Illinois Reemployment Bonus Experiment, and one with observational data, a female job training in Korea.

### 6.1 Illinois Reemployment Bonus Experiment

In this section we analyze data from the Illinois Reemployment Bonus Experiments, which is freely available at the W.E. Upjohn Institute for Employment Research. From mid-1984 to mid-1985, the Illinois Department of Employment Security conducted a social experiment to test the effectiveness of bonus offers in reducing the duration of insured unemployment. At the beginning of each claim, the experiment randomly divided newly unemployed people into three groups:

1. Job Search Incentive Group (JSI). The members of this group were told that they would qualify for a cash bonus of \$500, which was about four times the average weekly unemployment insurance benefits, if they found a full-time job within eleven weeks of benefits, and if they held that job for at least four months. 4816 claimants were assigned to this group.
2. Hiring Incentive Group (HI). The members of this group were told that their employer would qualify for a cash bonus of \$500 if the claimant found a full-time job within eleven weeks of benefits, and if they held that job for at least four months. 3963 claimants were assigned to this group.

3. Control Group. All claimants not assigned to the other groups. These members did not know that the experiment was taking place. 3952 individuals were assigned to this group.

Several studies including Woodbury and Spiegelman (1987), Meyer (1996) and Bijwaard and Ridder (2005) have analyzed the impact of the reemployment bonus on the unemployment duration measured by the number of weeks receiving unemployment insurance. It is important to emphasize that spells which reached the maximum amount of benefits or the state maximum number of weeks, 26, are censored, leading to censoring proportions of 38, 41 and 42 percent for the JSI, HI and the control group, respectively. Apart from the duration data, some information about claimants' background characteristics is also available: age, gender (Male =1), ethnicity (White =1), pre-unemployment earning and the weekly unemployment insurance benefits amount. For a complete description of the experiment and the available dataset, see Woodbury and Spiegelman (1987).

We start our analysis by plotting in Figure 1 the estimated overall treatment effect for the Job Search Incentive and the Hiring Incentive groups. From the discussion in Section 2.3, one can estimate  $\Upsilon(t)$  by

$$\hat{\Upsilon}(t) = \frac{1}{n} \sum_{i=1}^n \left( \frac{D_i \delta_i 1\{Q_i < \tau\}}{1 - \hat{G}_1^{KM}(Q_i-) \hat{p}(X_i)} - \frac{(1 - D_i) \delta_i 1\{Q_i < \tau\}}{(1 - \hat{G}_0^{KM}(Q_i-)) (1 - \hat{p}(X_i))} \right),$$

where, given the experimental design,  $\hat{p}(\cdot) = n^{-1} \sum_{i=1}^n D_i$ , which is numerically the same as the series logit estimator using a power series of order zero. Notice that both treatments seems to short the unemployment duration when compared to the control group, with the effects of the JSI group a bit larger than those of HI group.

We are focused on evaluating the effectiveness of the unemployment bonus in affecting the unemployment duration for all subgroups characterized by observable characteristics, and not just the overall effect as displayed in Figure 1. To this end, we perform our test for zero CDTE. We compare our results with the one using the semi-parametric Cox (1972) proportional hazard model.

Figure 1: Distributional treatment effects of Illinois reemployment bonus program on unemployment duration



The results of the tests are reported in Table 2. Using our nonparametric proposals, we reject the null hypothesis of zero CDTE at the 5% level for both treatment groups. Therefore, our tests suggest that the bonus experiment in Illinois were able to affect the unemployment duration. On the other hand, if one uses the proportional hazard model, one cannot reject the null of zero effect for all subpopulations in the hiring incentive group at the 5% level. In fact, by means of Grambsch and Therneau (1994)’s test, the proportionality assumption is rejected in the data at the 1% level. This illustrates how using parametric models to assess the existence of treatment effects might lead to “erroneous” conclusions.

One might be also interested is assessing the direction of the treatment effect, i.e., if the unemployment bonus program has led to a shorter or longer unemployment duration. Given the design of the Illinois experiment, it is plausible to assume that offering a reemployment bonus for the unemployed cannot lead to longer unemployment spells than in the control group, i.e., we might exclude the possibility that the treatment is “harmful”, that is, we can impose that

$$F_1(t|X = x) \geq F_0(t|X = x) \quad \forall (t, x) \in \mathcal{W}. \quad (27)$$

With (27), we can focus on single direction of departure of the null hypothesis of no distributional treatment effects for all subpopulations characterized by covariates. That is, with

the additional information in (27), we can test

$$H_0^{one} : F_1(t|X = x) = F_0(t|X = x) \quad \forall (t, x) \in \mathcal{W},$$

against

$$H_1^{one} : F_1(t|X = x) \geq F_0(t|X = x) \quad \forall (t, x) \in \mathcal{W}$$

*with strict inequality for some  $(t, x) \in \mathcal{W}$*

using the test statistic

$$KS_n^{one} = \sup_{(t,x) \in \mathcal{W}} \sqrt{n} \hat{I}_n(t, x). \quad (28)$$

Critical values are computed as described in Section 3.4.

As shown in Table 2, we reject the null of zero conditional treatment effects in favor of the one sided alternative that the treatment is non-negative (not-harmful) for all subpopulations, in both treatment groups. Even though we excluded the possibility of a “negative treatment effect” for the Illinois experiment, as a “robustness check”, we also consider the other one-sided alternative, i.e., the one in which the treatment is non-positive (not-helpful) for all subpopulations. In fact, we fail to reject our null hypothesis of zero conditional treatment effect for both treatment groups. Therefore, this evidence suggest that the bonus experiment has reduced the unemployment duration.

An important aspect of the Illinois Reemployment Bonus Experiment is that participation was not mandatory. Once claimants were assigned to the treatment groups, they were asked if they would like to participate in the demonstration or not. For those selected to the Job Search Incentive group, 84% agreed to participate, whereas just 65% of the Hiring Incentive group agreed to participate. This non-compliance issue may raise some selection bias issue. Therefore, the performed tests might be interpreted as tests for zero distributional “intention to treat” effects. Nonetheless, one may be willing to some extent disentangle the effects of participation and the effects of actual treatment. Using the random assignment as an instrumental variable for the actual participation in the demonstration, we adopt the

Table 2: Bootstrap p-values for conditional distributional treatment effects tests for the Illinois bonus experiment.

Intention to Treat		
Treatment Effect tests	Job Search Incentive	Hiring Incentive
Two sided	0.000	0.030
One sided - Not harmful	0.000	0.015
One sided - Not helpful	0.684	0.987
Proportional Hazard Model	0.000*	0.059*
Local Treatment Effects - Compliers		
Treatment Effect tests	Job Search Incentive	Hiring Incentive
Two sided	0.000	0.025
One sided - Not harmful	0.000	0.012
One sided - Not helpful	0.885	0.935

Note: 10,000 bootstrap replications. \* denotes p-value based on Gaussian approximation.

LTE framework described in Section 5.2, using a power series of order two<sup>2</sup>. The results for these tests are displayed in the second part of Table 2, and the conclusions of our tests are not changed. Therefore, we argue that indeed there is statistical evidence that the unemployment bonus experiment has helped their participants shorten their unemployment spell in Illinois.

## 6.2 Female job training in Korea.

Our method can also be used with observational data. Therefore, we analyze female job training data from the Department of Labor in South Korea in which the control group consist of unemployed claimants who chose to receive unemployment insurance instead of job training. This dataset is also used by Lee (2009). The data represents about 20% of the Korea population who became unemployed during January 1999 to the end of 1999 and either received job training or used unemployment insurance up to the end of 1999. All individuals were followed until the end of March 2000. Therefore, from the design of the data, the maximum unemployment duration is 450 days, that is,  $\tau_C = 450$ . Nonetheless, in the dataset, all observations with unemployment duration beyond 420 days are censored.

---

2. The results are robust to different choices of the number of power series considered in the estimation of the propensity score.

There are 9312 members individuals in the control group and 1554 in the treatment group. For a complete description of the dataset and the characteristics of the job training program, see Lee and Lee (2005).

In addition to the unemployment duration information, we use as covariates informations on the individual characteristics such as the number of days that the woman worked at her last workplace, education (completed high school=1), age in years, and four ex-job categories (1-executive, professional or semiprofessional; 2-clerical; 3-service or sales; 4-mechanic, assembler, operator and menial labour). In the dataset, the proportion of censoring is around 70% for both treated and control groups. Notice that the duration for the treated group includes the duration of job training, which average duration was about 3 months.

As is usual in the policy evaluation literature, we first estimate the unconditional average trimmed treatment effect,  $\mathbb{E}[Y_1 1\{Y_1 < \tau_C\} - Y_0 1\{Y_0 < \tau_C\}]$ , by

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{D_i \delta_i Q_i}{1 - \hat{G}_1^{KM}(Q_i-) \hat{p}(X_i)} - \frac{(1 - D_i) \delta_i Q_i}{(1 - \hat{G}_0^{KM}(Q_i-)) (1 - \hat{p}(X_i))} \right). \quad (29)$$

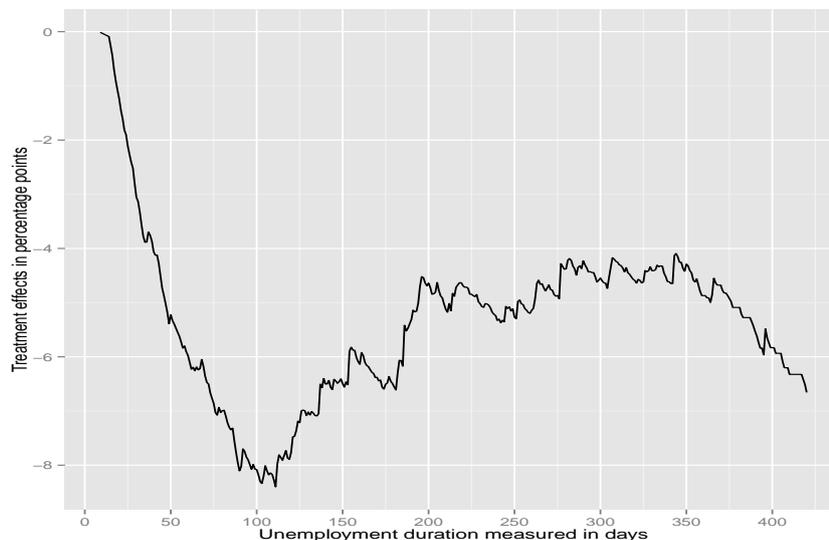
The ATE point estimate is approximately -5 days, i.e. the job training had reduced the overall unemployment duration by 5 days. Nonetheless, following an analogous procedure as describe in Section 3, we are not able to reject the null hypothesis that the ATE is equal to zero at the conventional levels. Therefore, looking at the *unconditional* ATE, one may argue that the unemployment duration for those who receive unemployment insurance and those who received job training are the same.

Next, we consider the unconditional distributional treatment effect. Figure 2 plots  $\hat{\Upsilon}(t)$ , but with the propensity score  $p(\cdot)$  estimated with the series logit estimator using a power series of order two<sup>3</sup>. From the plot, one may argue that job training leads to an increase in the unemployment duration of female Koreans. Indeed, we reject the null hypothesis that the unconditional DTE is zero. From the results of the unconditional ATE and DTE tests, it seems that the job training has had an effect at the unemployment distribution at some point other than the average.

---

3. Our results are not sensitive to different choices of the number of power series included.

Figure 2: Distributional treatment effects of job training on female unemployment duration in South Korea



In order to analyze if this conclusion holds true after conditioning on a vector of individual characteristics, we apply our tests for both zero conditional average treatment effect and for zero conditional distributional treatment effect, using age, last firm employment days, education level and job categories as covariates. For comparison, we also apply Lee (2009)’s test for zero CDTE based on a ‘two-sample’ covariate-matching procedure. To avoid dimensionality problems, we only consider matching on the propensity score (estimated with a probit), with bandwidth equal to  $0.62n^{-1/5}$ , and the bi-weight kernel  $K(z) = (15/16)(1 - z^2)^2 1\{|z| < 1\}$ .

The results are presented in Table 3. Using both Lee (2009)’s and our proposal, we find evidence of treatment effect has an effect on the distributional of unemployment duration at the 5% level. Furthermore, we reject the null hypothesis of zero conditional average treatment effect at the 5% level, which is in contrast to the unconditional case. Hence, we conclude that, after conditioning on a vector of observables, we are able to point out the direction of the departure of the null hypothesis (1). This illustrates the complementarity of our tests for zero CDTE and CATE, and the additional information these tests can provide when compared to their unconditional counterparts.

Table 3: Tests for the Korean job training data . Bootstrap p-values

Treatment Effect tests	Bootstrap p-value
Cond. Dist. Treat. Effect	0.000
Cond. Aver. Treat. Effect	0.004
Cond. Dist. Treat. Effect - Lee(2009)	0.000 *

Note: 10,000 bootstrap replications. \* denotes p-value based on Gaussian approximation.

## 7 Conclusion and suggestions for further research

In this paper we proposed different nonparametric tests for treatment effects when the outcome of interest is censored. Once we transform our conditional moment restrictions into an infinite number of moments, we characterize our tests statistics as Kaplan-Meier integrals that can be easily estimated from the observed data. Our tests have asymptotically correct size, are able to detect local alternatives converging to the null hypothesis of interest at the parametric rate  $n^{-1/2}$ , and are consistent against fixed alternatives. Our simulation study provide evidence that our tests have good finite sample properties. We provide two empirical applications to demonstrate that our tests can be useful in practice.

Our results can be extended to other situations of practical interest. For instance, an interesting extension of our results consists of testing conditional stochastic dominance when the outcome of interest is a duration. In the context of fully observed data, conditional stochastic dominance has recently attracted a lot of interest. See, for example, Lee and Whang (2009), Delgado and Escanciano (2013), and Lee et al. (2013). Adopting either Delgado and Escanciano (2013)'s or Andrews and Shi (2013, 2014)'s approach, one can extend our proposal to the stochastic dominance analysis to censored outcomes.

Another important extension would be to allow the covariates distribution to be different in the treatment and control groups by introducing covariate-matching techniques. With these techniques, the use of smooth estimators cannot be avoided. In particular, proposals by Cabus (1998), Neumeyer and Dette (2003), and Srihera and Stute (2010) designed for testing the equality of nonparametric regression curves in a two-sample context with fully observed data, can be adapted to handle randomly censored outcomes if one use Kaplan-

Meier integrals as in this article. For a related approach with censored outcomes, see Pardo-Fernandez and Van Keilegom (2006) and Lee (2009). A detailed analysis of these extensions is beyond the scope of this article and is deferred to future work.

## A Appendix A: Proofs of main Theorems

In this appendix, we prove our main results. Before proving the main results of the article, we first introduce some notation. For a generic set  $\mathcal{G}$ , let  $l^\infty(\mathcal{G})$  be the Banach space of all uniformly bounded real functions on  $\mathcal{G}$  equipped with the uniform metric  $\|f\|_{\mathcal{G}} \equiv \sup_{z \in \mathcal{G}} |f(z)|$ . We consider convergence in distribution of empirical processes in the metric space  $(l^\infty(\mathcal{G}), \|f\|_{\mathcal{G}})$  in the sense of J. Hoffman-Jørgensen (see, e.g. , van der Vaart and Wellner (1996)). For any generic Euclidean random vector  $\xi$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $\chi_\xi$  denotes its state space and  $P_\xi$  denotes its induced probability measure with corresponding distribution function  $F_\xi(\cdot) = P_\xi(-\infty, \cdot]$ . Given *iid* observations  $\{\xi_i\}_{i=1}^n$  of  $\xi$ ,  $\mathbb{P}_{\xi_n} f \equiv n^{-1} \sum_{i=1}^n f(\xi_i)$ . Let  $F_{\xi_n}(\cdot) \equiv \mathbb{P}_{\xi_n}(-\infty, \cdot]$  be the corresponding empirical CDF. Likewise, the expectation is denoted by  $P_\xi f \equiv \int f dP$ . The empirical process evaluated at  $f$  is  $\mathbb{G}_{\xi_n} f$  with  $\mathbb{G}_{\xi_n} \equiv \sqrt{n}(\mathbb{P}_{\xi_n} - P_{\xi_n})$ . Let  $\|\cdot\|_{2,P}$  be the  $L_2(P)$  norm, that is,  $\|f\|_{2,P}^2 = \int f^2 dP$ . When  $P$  is clear from the context, we simply write  $\|\cdot\|_2 \equiv \|\cdot\|_{2,P}$ . Let  $|\cdot|$  denote the Euclidean norm, that is,  $|A|^2 = A'A$ . For a measurable class of functions  $\mathcal{G}$  from  $\chi_Z$  to  $\mathbb{R}$ , let  $\|\cdot\|$  be a pseudo-norm on  $\mathcal{G}$ , that is, a norm except for the property  $\|f\| = 0$  does not imply  $f = 0$ . Let  $N(\varepsilon, \mathcal{G}, \|\cdot\|)$  be the covering number with respect to  $\|\cdot\|$  needed to cover  $\mathcal{G}$ . Given two functions  $l, u \in \mathcal{G}$ , the bracket  $[l, u]$  is the set of functions  $f \in \mathcal{G}$  such that  $l \leq f \leq u$ . An  $\varepsilon$ -bracket with respect to  $\|\cdot\|$  is a bracket  $[l, u]$  with  $\|l - u\| \leq \varepsilon$ . The covering number with bracketing  $N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|)$  is the minimal number of  $\varepsilon$ -brackets with respect to  $\|\cdot\|$  needed to cover  $\mathcal{G}$ . Define  $\mathcal{S} \equiv (-\infty, \tau_C) \times \chi_X$ . Throughout the appendix, denote  $\mathcal{C}$  as a generic constant that may change from expression to expression.

First, we present the proof of the identification result in Proposition 1.

**Proof of Proposition 1:** By Assumptions 2-4 and the law of iterated expectations, we have, for  $t < \tau_C$

$$\begin{aligned}
& \mathbb{E} \left[ \frac{D\delta 1\{Q \leq t\}}{(1 - G_1(Q-))p(X)} \middle| X \right] \\
&= \mathbb{E} \left[ \frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))p(X)} \mathbb{E}[D\delta_1 | X, Y_1] \middle| X \right] \\
&= \mathbb{E} \left[ \frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))p(X)} \mathbb{E}[D | X] \mathbb{E}[\delta_1 | X, Y_1] \middle| X \right] \\
&= \mathbb{E} \left[ \frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))p(X)} p(X) (1 - G_1(Y_1)) \middle| X \right] \\
&= \mathbb{E}[1\{Y_1 \leq t\} | X],
\end{aligned}$$

where the first equality follows from the law of iterated expectations, the second from Assumption 2, and the third from Assumption 4. Assumption 3 guarantees that the expectation is well defined.

Adopting the analogous steps for the control group,

$$\begin{aligned} & \mathbb{E} \left[ \frac{(1-D)\delta 1\{Q \leq t\}}{(1-G_0(Q-))(1-p(X))} \middle| X \right] \\ &= \mathbb{E} [1\{Y_0 \leq t\} | X]. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{D\delta 1\{Q \leq t\}}{(1-G_1(Q-))p(X)} - \frac{(1-D)\delta 1\{Q \leq t\}}{(1-G_0(Q-))(1-p(X))} \right) \middle| X \right] \\ &= \Upsilon(t, x) \end{aligned}$$

concluding the proof. ■

Next, we state an auxiliary result from the empirical process literature. Define the generic class of measurable functions  $\mathcal{G} \equiv \{Z \rightarrow m(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ , where  $\Theta$  and  $\mathcal{H}$  are endowed with the pseudo-norms  $|\cdot|_{\Theta}$  and  $|\cdot|_{\mathcal{H}}$ . The following result is part of Theorem 3 in Chen et al. (2003).

**Lemma A.1** *Assume that for all  $(\theta_0, h_0) \in \Theta \times \mathcal{H}$ ,  $m(Z, \theta, h)$  is locally uniformly  $L_2(P)$  continuous, in the sense that*

$$E \left[ \sup_{\theta: |\theta - \theta_0|_{\Theta} < \delta, h: |h - h_0|_{\mathcal{H}} < \delta} |m(Z, \theta, h) - m(Z, \theta_0, h_0)|^2 \leq C\delta^s \right]$$

for all sufficiently small  $\delta > 0$  and some constant  $s \in (0, 2]$ . Then,

$$\begin{aligned} N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|_2) &\leq N \left( \left( \frac{\varepsilon}{2C} \right)^{2/s}, \Theta, \|\cdot\|_{\Theta} \right) \\ &\quad \times N \left( \left( \frac{\varepsilon}{2C} \right)^{2/s}, \mathcal{H}, \|\cdot\|_{\mathcal{H}} \right). \end{aligned}$$

Before we introduce the proofs of our main theorems, we prove two useful lemmas that are crucial to the derivation of our result. Recall that, for a given  $(t, x) \in \mathcal{W}$ ,

$$\begin{aligned} \xi_1(\bar{y}, \bar{x}, \bar{z}; t, x) &= \bar{z}(1-p(\bar{x}))1\{\bar{y} \leq t\}1\{\bar{x} \leq x\}, \\ \xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) &= (1-\bar{z})p(\bar{x})1\{\bar{y} \leq t\}1\{\bar{x} \leq x\}. \end{aligned}$$

where  $p(\cdot)$  is the true propensity score. Define the infeasible estimator

$$\begin{aligned} \bar{I}(t, x) &= \int \xi_1(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}^{KM}(d\bar{y}, d\bar{x}, \bar{z} = 1) \\ &\quad - \int \xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) \hat{F}^{KM}(d\bar{y}, d\bar{x}, \bar{z} = 0), \end{aligned} \tag{A.1}$$

the analogous of (10) but with the true propensity score.

**Lemma A.2** *Under Assumptions 1-5,*

$$\sup_{(t,x) \in \mathcal{S}} \left| \bar{I}(t,x) - \frac{1}{n} \sum_{i=1}^n \eta_i(t,x) \right| = o(n^{-1/2})$$

**Proof** To prove this lemma, it suffices to apply Theorem 1 of Sellero et al. (2005). Toward this goal, define the following class of real-valued measurable functions on  $\chi_Y \times \chi_X \times \{0, 1\}$

$$\mathcal{G}_1 \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_1(\bar{y}, \bar{x}, \bar{z}; t, x) : (t, x) \in \mathcal{S}\}. \quad (\text{A.2})$$

Notice that  $\mathcal{G}_1$  is a VC-subgraph class of functions with VC index smaller or equal than  $k + 2$  and admits the envelope  $\Phi(\bar{\omega}, \bar{x}, \bar{z}) = 1$  that satisfies the required moment conditions of Theorem 1 of Sellero et al. (2005). The same holds for the class of functions

$$\mathcal{G}_2 \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) : (t, x) \in \mathcal{S}\}. \quad (\text{A.3})$$

Hence, we have

$$\bar{I}(t,x) = \frac{1}{n} \sum_{i=1}^n \eta_i(t,x) + R_n(t,x) \quad (\text{A.4})$$

where

$$\sup_{(t,x) \in \mathcal{S}} |R_n(t,x)| = O\left(\frac{\ln^3 n}{n}\right) \text{ a.s.},$$

concluding our proof. ■

In the next lemma we focus on the treated group. The result for the control group is analogous.

**Lemma A.3** *Under Assumptions 1-8, we have, uniformly in  $(t, x) \in \mathcal{S}$*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{1 - \hat{G}_1^{KM}(Q_{i-})} - \frac{1}{1 - G_1(Q_{i-})} \right) \delta_i D_i 1\{Q_i \leq t\} 1\{X_i \leq x\} (\hat{p}(X_i) - p(X_i)) \\ & = o_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

where  $G_1(t-) \equiv \mathbb{P}(C_1 < t)$ .

**Proof** Denote

$$Z_1(t) = \frac{\hat{G}_1^{KM}(t-) - G_1(t-)}{1 - \hat{G}_1^{KM}(t-)}.$$

Now, one can write

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \hat{G}_1^{KM}(Q_{i-})} \delta_i D_i 1\{Q_i \leq t\} 1\{X_i \leq x\} (\hat{p}(X_i) - p(X_i)) \\
&= \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - G_1(Q_{i-})} \delta_i D_i 1\{Q_i \leq t\} 1\{X_i \leq x\} (\hat{p}(X_i) - p(X_i)) \\
&+ \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - G_1(Q_{i-})} Z_1(Q_i) \delta_i D_i 1\{Q_i \leq t\} 1\{X_i \leq x\} (\hat{p}(X_i) - p(X_i)) \\
&\equiv A_1(t, x) + A_2(t, x).
\end{aligned}$$

It suffices to show that, uniformly in  $(t, x) \in \mathcal{W}$ ,  $\sqrt{n}A_2(\cdot, \cdot) = o_{\mathbb{P}}(1)$ . First, rewrite  $\sqrt{n}A_2(\cdot, \cdot)$  as

$$\begin{aligned}
\sqrt{n}A_2(t, x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \left( \hat{G}_1^{KM}(Q_{i-}) - G_1(Q_{i-}) \right) \frac{1 - G_1(Q_{i-})}{1 - \hat{G}_1^{KM}(Q_{i-}) (1 - G_1(Q_{i-}))^2} \\
&\quad \times D_i \delta_i 1\{Q_i \leq t\} 1\{X_i \leq x\} (\hat{p}(X_i) - p(X_i)).
\end{aligned}$$

We have

$$\begin{aligned}
\sqrt{n} \sup_{(t,x) \in \mathcal{W}} |A_2(t, x)| &\leq \mathcal{C} \sqrt{n} \sup_t \left| \hat{G}_1^{KM}(t) - G_1(t) \right| \times \sup_t \left| \frac{1 - G_1(t)}{1 - \hat{G}_1^{KM}(t)} \right| \times \sup_x |(\hat{p}(x) - p(x))| \\
&= O_{\mathbb{P}}(1) \times O_{\mathbb{P}}(1) \times \left[ O_{\mathbb{P}} \left( \sqrt{\frac{L^3}{n}} \right) + O(L^{-\frac{s}{2k}+1}) \right]
\end{aligned}$$

in which the last step follows from

$$\sqrt{n} \sup_t \left| \hat{G}_1^{KM}(t) - G_1(t) \right| = O_{\mathbb{P}}(1), \tag{A.5}$$

$$\sup_t \left| \frac{1 - G_1(t)}{1 - \hat{G}_1^{KM}(t)} \right| = O_{\mathbb{P}}(1), \tag{A.6}$$

$$\sup_x |(\hat{p}_n(x) - p(x))| = O_{\mathbb{P}} \left( \sqrt{\frac{L^3}{n}} \right) + O(L^{-\frac{s}{2k}+1}) \tag{A.7}$$

see Gill (1983) for (A.5), Zhou (1992) for (A.6), and Hirano et al. (2003) for (A.7). Taking  $L = a \cdot N^v$  as in Assumption 8,

$$\begin{aligned}
O_{\mathbb{P}} \left( \sqrt{\frac{L^3}{n}} \right) + O_{\mathbb{P}}(L^{-\frac{s}{2k}+1}) &= O_{\mathbb{P}} \left( n^{\frac{3v-1}{2}} \right) + O \left( n^{-(\frac{s}{2k}+1)v} \right) \\
&= o_{\mathbb{P}}(1)
\end{aligned}$$

if  $v < 1/3$  and  $s/k > 2$ . From Assumptions 7 and 8, these two conditions are fulfilled, concluding our proof. ■

Now we are ready to proceed with the proofs of our main results.

**Proof of Lemma 1:** Notice that

$$\begin{aligned} \hat{I}(t, x) &= \bar{I}(t, x) \\ &+ \frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t, x) - \xi_1(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t, x) \right) \\ &- \frac{n_0}{n} \sum_{l=1}^{n_0} W_{0,l:n_0} \left( \hat{\xi}_0(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x) - \xi_0(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x) \right) \end{aligned} \quad (\text{A.8})$$

First, by Lemma A.2, we have that, uniformly in  $(t, x) \in \mathcal{S}$ ,

$$\bar{I}(t, x) = \frac{1}{n} \sum_{i=1}^n \eta_i(t, x) + o_{\mathbb{P}}(n^{-1/2}). \quad (\text{A.9})$$

We now focus on the second term of (A.8). Our goal is to show that

$$\begin{aligned} \sup_{(t,x) \in \mathcal{S}} \left| \frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) - \xi_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) \right) \right. \\ \left. - \frac{1}{n} \sum_{i=1}^n F_1(t|X_i) p(X_i) 1\{X_i \leq x\} (D_i - p(X_i)) \right| = o_{\mathbb{P}}(n^{-1/2}) \end{aligned} \quad (\text{A.10})$$

As discussed in Section 3, we have that

$$W_{1,i:n_1} = \frac{\delta_{[i:n_1]}}{n_1} \frac{1}{1 - \hat{G}_1^{KM}(Q_{i:n_1}-)},$$

where  $\hat{G}_1^{KM}$  is the Kaplan-Meier estimator of  $G_1$ . By Lemma A.3, we have that, uniformly in  $(t, x) \in \mathcal{S}$

$$\begin{aligned} &\frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) - \xi_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{D_i \delta_i}{1 - G_1(Q_i-)} \left( \hat{\xi}_1(Q_i, X_i, D_i; t, x) - \xi_1(Q_i, X_i, D_i; t, x) \right) + o_{\mathbb{P}}(n^{-1/2}), \end{aligned} \quad (\text{A.11})$$

that is, that there is no estimation effect due to the replacing  $G_1$  by estimation of  $\hat{G}_1^{KM}$  in the second term of (A.8).

By adding and subtracting a number of terms, we have that (A.11) is equal to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \frac{D\delta_i}{1 - G_1(Q_i -)} \left( \hat{\xi}_1(Q_i, X_i, D_i; t, x) - \xi_1(Q_i, X_i, D_i; t, x) \right) - \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) (\hat{p}_n(\bar{x}) - p(\bar{x})) \mathbb{P}(d\bar{x}) \right] \quad (\text{A.12})$$

$$+ \left[ \sqrt{n} \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) (\hat{p}_n(\bar{x}) - p(\bar{x})) \mathbb{P}(d\bar{x}) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{u}_n(X_i) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X_i)(1 - p_L(X_i))}} \right] \quad (\text{A.13})$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{u}_n(X_i) - u_n(X_i)) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X_i)(1 - p_L(X_i))}} \quad (\text{A.14})$$

$$+ \frac{1}{\sqrt{n}} \left( \sum_{i=1}^n u_n(X_i) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X_i)(1 - p_L(X_i))}} - u(X_i) \frac{(D_i - p(X_i))}{\sqrt{p(X_i)(1 - p(X_i))}} \right) \quad (\text{A.15})$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^n F_1(t|X_i) 1\{X_1 \leq x\} p(X_i) (D_i - p(X_i)) \quad (\text{A.16})$$

where

$$\begin{aligned} \tilde{u}_n(z) &= \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) \mathcal{L}'(R^L(\bar{x})' \tilde{\pi}_L) R^L(\bar{x})' \mathbb{P}(d\bar{x}) \tilde{\Sigma}_L^{-1} \sqrt{p_L(X_i)(1 - p_L(X_i))} R^L(z), \\ u_n(z) &= \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) \mathcal{L}'(R^L(\bar{x})' \pi_L) R^L(\bar{x})' \mathbb{P}(d\bar{x}) \Sigma_L^{-1} \sqrt{p_L(X_i)(1 - p_L(X_i))} R^L(z), \\ u(z) &= F_1(t|z) 1\{z \leq x\} p(z) \sqrt{p(z)(1 - p(z))}, \end{aligned}$$

with

$$\begin{aligned} \Sigma_L^{-1} &= \mathbb{E} \left[ R^L(X) R^L(X)' \mathcal{L}'(R^L(\bar{x})' \pi_L) \right] \\ \tilde{\Sigma}_L^{-1} &= \frac{1}{n} \sum_{i=1}^n R^L(X_i) R^L(X_i)' \mathcal{L}'(R^L(\bar{x})' \tilde{\pi}_L), \end{aligned}$$

and  $\tilde{\pi}_L$  between  $\hat{\pi}_L$  and  $\pi_L$ .

By the same arguments as in the Addendum of ?<sup>4</sup>, we have that, uniformly in  $(t, x) \in \mathcal{S}$ ,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{D_i \delta_i}{1 - G_1(Q_i -)} \left( \hat{\xi}_1(Q_i, X_i, D_i; t, x) - \xi_1(Q_i, X_i, D_i; t, x) \right) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^n F_1(t|X_i) p(X_i) 1\{X_1 \leq x\} (D_i - p(X_i)) \\
& = \left[ O_{\mathbb{P}}(L^{-\frac{s}{2k}+1}) + O_{\mathbb{P}}(L^2 n^{-\frac{1}{2}}) \right] + O_{\mathbb{P}}(\sqrt{n} L^{-\frac{s}{2k}+1}) \\
& \quad + O_{\mathbb{P}}(n^{-\frac{1}{2}} L^{\frac{11}{2}}) + O_{\mathbb{P}}\left(\max\left(L^{1-\frac{s}{2k}}, L^{-\frac{1}{2k}}\right)\right) \\
& = o_{\mathbb{P}}(1)
\end{aligned}$$

under Assumptions 7 and 8. Therefore, by the above results we obtain that

$$\begin{aligned}
& \sup_{(t,x) \in \mathcal{S}} \left| \frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) - \xi_1(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x) \right) \right. \quad (\text{A.17}) \\
& \quad \left. - \frac{1}{n} \sum_{i=1}^n F_1(t|X_i) 1\{X_i \leq x\} p(X_i) (D_i - p(X_i)) \right| \\
& = o_{\mathbb{P}}(n^{-1/2})
\end{aligned}$$

as desired.

By applying the same arguments, we have that

$$\begin{aligned}
& \sup_{(t,x) \in \mathcal{S}} \left| \frac{n_0}{n} \sum_{l=1}^{n_0} W_{0,l:n_0} \left( \hat{\xi}_0(Q_{l:n_0}, X_{[l:n_0]}, D_{[l:n_0]}; t, x) - \xi_0(Q_{l:n_0}, X_{[l:n_0]}, D_{[l:n_0]}; t, x) \right) \right. \quad (\text{A.18}) \\
& \quad \left. - \frac{1}{n} \sum_{i=1}^n F_0(t|X_i) (1 - p(X_i)) 1\{X_1 \leq x\} (D_i - p(X_i)) \right| = o_{\mathbb{P}}(n^{-1/2}).
\end{aligned}$$

Combining (A.9), (A.17) and (A.18), we have that, uniformly in  $(t, x) \in \mathcal{S}$ ,

$$\begin{aligned}
& \hat{I}(t, x) - I(t, x) \\
& = \frac{1}{n} \sum_{i=1}^n [(\eta_i(t, x) - I(t, x)) - \alpha(X_i; t, x) (D_i - p(X_i))] + o_{\mathbb{P}}(n^{-1/2})
\end{aligned}$$

where

$$\alpha(X_i; t, x) = [F_1(t|X_i) p(X_i) + F_0(t|X_i) (1 - p(X_i))] 1\{X_i \leq x\}$$

concluding our proof. ■

**Proof of Theorem 1:** From the asymptotic representation of Lemma 1, it suffices to prove the convergence of the dominant term. To this end, define the class of real-valued

---

4. The step-by-step procedure is available upon request.

measurable functions on  $\chi_Y \times \chi_X \times \{0, 1\} \times \{0, 1\}$

$$\mathcal{F} = \{(\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}) \rightarrow \varphi_{(t,x)} \equiv \eta(\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}; t, x) - (F_1(t|\bar{x})p(\bar{x}) + F_0(t|\bar{x})(1 - p(\bar{x})))1\{\bar{x} \leq x\}(\bar{z} - p(\bar{x})) : (t, x) \in \mathcal{S}\}$$

where  $\eta(\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}; t, x)$  is defined as in (15)

Our goal is to show that class of functions  $\mathcal{F}$  is Donsker. By Theorem 2.10.6 in van der Vaart and Wellner (1996) it suffices to show that the classes of functions  $\mathcal{G}_1$  and  $\mathcal{G}_2$ , as defined in (A.2) and (A.3), and for  $j = \{0, 1\}$ ,  $\{\gamma_{j,0}(\cdot)\}, \{\delta\}, \{\gamma_{j,1}(\cdot)\}, \{\gamma_{j,2}(\cdot)\}, \{F_j(t, x|\cdot)\}, \{(D - p(\cdot))\}$  are Donsker.

For  $j = \{0, 1\}$ , define the class of real-valued measurable functions on  $\chi_X$

$$\mathcal{G}_{3,j} \equiv \{\bar{x} \rightarrow \phi_2(\bar{x}) \equiv F_j(t|\bar{x}) : t \in \mathcal{S}\}. \quad (\text{A.19})$$

Now, notice that both  $\mathcal{G}_1$ ,  $\mathcal{G}_2$  and  $\mathcal{G}_{3,j}$  are VC-Class with square integrable envelope functions. Therefore, by Theorem 2.6.8 in van der Vaart and Wellner (1996), these classes of functions are Donsker. The functions  $\gamma_{0,j}$ ,  $(D - p(\cdot))$  and  $\delta$  does not depend on  $t$  nor on  $x$  and so they are clearly Donsker.

We next consider  $\gamma_{j,1}$ . For  $j = \{0, 1\}$ , define the classes of real-valued measurable functions

$$\mathcal{F}_{1,j} = \{\omega \in [-\infty, t_H) \rightarrow \gamma_{j,1}(\omega) \equiv \frac{1}{1 - H_j(\omega)} \int 1\{\omega < \bar{\omega}\} \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x) \gamma_{j,0}(\bar{\omega}) H_{j,11}(d\bar{\omega}, d\bar{x}) : (t, x) \in \mathcal{S}\}$$

$$\mathcal{F}_{2,j} = \{\omega \in [-\infty, t_H) \rightarrow \gamma_{j,2}(\omega) \equiv \int \frac{1\{\bar{v} < \omega, \bar{v} < \bar{\omega}\} \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x)}{[1 - H_j(\bar{v})]^2} \gamma_{j,0}(\bar{\omega}) H_{j,0}(d\bar{v}) H_{j,11}(d\bar{\omega}, d\bar{x}) : (t, x) \in \mathcal{S}\}$$

In order to prove that these classes of functions are Donsker, by Theorem 2.5.6 of van der Vaart and Wellner (1996), it suffices to show that, for  $i = 1, 2$ ,

$$\int_0^\infty \sqrt{\ln N_{[\cdot]}(\varepsilon, \mathcal{F}_{i,j}, L_2(P))} d\varepsilon < \infty \quad (\text{A.20})$$

where  $P$  is the probability measure corresponding to the joint distribution of  $(Q, \delta, D, X)$ , and  $L_2(P)$  is the  $L_2$ -norm. Notice that both  $\mathcal{F}_{1,j}$  and  $\mathcal{F}_{2,j}$  are classes of monotone bounded functions. Therefore, by Theorem 2.7.5 in van der Vaart and Wellner (1996), we have that, for a fixed  $\varepsilon > 0$  and  $i = 1, 2$ ,  $\ln N_{[\cdot]}(\varepsilon, \mathcal{F}_{i,j}, L_2(P)) \leq K\varepsilon^{-1}$ , where  $K$  is an arbitrary constant. Hence, for  $i = 1, 2$ , the integral in (A.20) is finite, and the classes of functions  $\mathcal{F}_{1,j}$  and  $\mathcal{F}_{2,j}$ ,  $j = \{0, 1\}$ , are Donsker.

We have just shown that  $\mathcal{F}$  is Donsker, that is, we have proved that

$$\sqrt{n}(\hat{I} - I) \Rightarrow C_\infty$$

where  $C_\infty$  is a tight Gaussian process in  $l^\infty(\mathcal{S})$  with zero mean and covariance function given by (20). Since under  $H_0$ ,  $I(t, x) = 0 \forall (t, x) \in \mathcal{W} \subseteq \mathcal{S}$ , the proof is completed. ■

**Proof of Theorem 2:** Notice that we can always write

$$\begin{aligned}\sqrt{n}\hat{I}(t, x) &= \sqrt{n}(\hat{I} - I)(t, x) + \sqrt{n}I(t, x) \\ &= D_{1,n}(t, x) + D_{2,n}(t, x).\end{aligned}$$

From the proof of Theorem 1, we have that

$$\sqrt{n}(\hat{I} - I) \Rightarrow C_\infty,$$

and therefore  $D_{1,n}(t, x) = O_{\mathbb{P}}(1)$ . On the other hand, under the alternative  $I(t, x) \neq 0$  for some  $(t, x)$ . Therefore  $D_{2,n}(t, x) = O_{\mathbb{P}}(n^{1/2})$ . Hence, under  $H_1$ ,

$$\sqrt{n} \sup_{(t,x) \in \mathcal{W}} |\hat{I}(t, x)| \xrightarrow{p} \infty,$$

Since under  $H_0$ ,  $I(t, x) = 0$  for all  $(t, x)$ ,  $KS_n = O_{\mathbb{P}}(1)$ , and therefore  $c_\alpha^{KS} = O(1)$  almost surely, we conclude that

$$\lim_{n \rightarrow \infty} P \{KS_n > c_\alpha^{KS}\} = 1.$$

Analogously, we have that

$$\lim_{n \rightarrow \infty} P \{CvM_n > c_\alpha^{CvM}\} = 1.$$

■

**Proof of Theorem 3:** As in the proof of Theorem 2, we can always write

$$\begin{aligned}\sqrt{n}\hat{I}(t, x) &= \sqrt{n}(\hat{I} - I)(t, x) + \sqrt{n}I(t, x) \\ &= D_{1,n}(t, x) + D_{2,n}(t, x)\end{aligned}$$

From the proof of Theorem 1, we have that

$$\sqrt{n}(\hat{I} - I) \Rightarrow C_\infty,$$

and therefore  $D_{1,n}(t, x) = O_{\mathbb{P}}(1)$ . On the other hand, under the local alternatives of the type  $H_{1,n}$ ,  $\sqrt{n}I(t, x) = \mathbb{E}[h(t, x)(p(X)(1 - p(X)))1\{X \leq x\}] = O_{\mathbb{P}}(1)$ . Hence, under  $H_{1,n}$ ,

$$\sqrt{n}\hat{I}(t, x) \Rightarrow C_\infty + R(t, x)$$

in  $l^\infty(\mathcal{W})$ . ■

Before we proceed with the proof of Theorem 4, we prove the uniform consistency of our estimator for

$$\begin{aligned}\alpha(X; t, x) &= \left( \mathbb{E} \left[ \frac{D\delta 1\{Q \leq t\}}{1 - G_1(Q-)} \middle| X \right] - \mathbb{E} \left[ \frac{(1 - D)\delta 1\{Q \leq t\}}{1 - G_0(Q-)} \middle| X \right] \right) 1\{X \leq \cdot\} \\ &= [F_1(t|X)p(X) - F_0(t|X)(1 - p(X))] 1\{X \leq \cdot\}\end{aligned}$$

To this end, it suffices to show that

$$\sup_{(t,\bar{x}) \in \mathcal{S}} |\hat{\alpha}_1^{KM}(\bar{x}; t) - F_1(t|\bar{x})p(\bar{x})| = o_{\mathbb{P}}(1),$$

where

$$\hat{\alpha}_1^{KM}(\bar{x}; t) = \left( \frac{1}{n} \sum_{i=1}^n \frac{D_i \delta_i 1\{Q_i \leq t\} R^L(X_i)}{1 - \hat{G}_1^{KM}(Q_{i-})} \right)' \left( \frac{1}{n} \sum_{i=1}^n R^L(X_i) R^L(X_i)' \right)^{-1} R^L(\bar{x})$$

The analogous result applies to the other expectation

**Lemma A.4** *Suppose Assumptions 1-8 hold. Additionally, assume that  $F_1(\cdot, \cdot|X)$  is continuously differentiable of order  $m \geq k$ , where  $k$  is the dimension of  $X$ . Then,*

$$\sup_{(t,\bar{x}) \in \mathcal{S}} |\hat{\alpha}_1^{KM}(\bar{x}; t) - F_1(t|\bar{x})p(\bar{x})| = o_{\mathbb{P}}(1)$$

**Proof** For a matrix  $A$ , let  $\|A\|$  denote the matrix norm of  $A$  such that  $\|A\| = \sqrt{\text{tr}(A'A)}$ . Define

$$\begin{aligned} \Phi_L(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i D_i 1\{Q_i \leq t\} \gamma_{1,0}(Q_i) R^L(X_i), \\ \Phi_L^{KM}(t) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i D_i 1\{Q_i \leq t\} R^L(X_i)}{1 - \hat{G}_1^{KM}(Q_{i-})}, \\ \zeta_L &= \frac{1}{n} \sum_{i=1}^n R^L(X_i) R^L(X_i)'. \end{aligned}$$

Notice that

$$\hat{\alpha}_1^{KM}(t|\bar{x}) = \Phi_L^{KM}(t)' \zeta_L^{-1} R^L(\bar{x})$$

From Theorem 1 of Stute (1993), we have that  $\Phi_L^{KM}(t) = \Phi_L(t)$  *a.s.*. Given that the conditional variance of  $\delta_i 1\{Q \leq \cdot\} \gamma_0(Q_i)$  conditional on  $X$  is bounded, the uniform bound in Newey (1997) for power series estimators applies:

$$\sup_{(t,\bar{x}) \in \mathcal{S}} \left| \Phi_L(t)' \zeta_L^{-1} R^L(\bar{x}) - F_1(t|\bar{x})p(\bar{x}) \right| \leq \mathcal{C} \left( L^{\frac{3}{2}} n^{-\frac{1}{2}} + L^{1-\frac{m}{k}} \right)$$

where  $m$  is the number of continuous derivatives of  $F(\cdot|\bar{x})$ .

Taking  $L = a \cdot N^v$  as in Assumption 8, and from the results above, we have that

$$\sup_{(t,\bar{x}) \in \mathcal{S}} |\hat{\alpha}_1^{KM}(\bar{x}; t) - F_1(t|\bar{x})p(\bar{x})| = o_{\mathbb{P}}(1)$$

if  $v < 1/3$ , and  $m \geq k$ . Given that these conditions are fulfilled, we conclude our proof. ■

Next, we proceed with the proof of Theorem 4.

**Proof of Theorem 4:** For  $j \in \{0, 1\}$ , denote

$$\hat{\eta}_{j,i}(t, x) = \hat{\xi}_j(Q_i, X_i, D_i; t, x) \hat{\gamma}_{j,0}(Q_{j,i}) \delta_{j,i} + \hat{\gamma}_{j,1}(Q_{j,i}) (1 - \delta_{j,i}) - \hat{\gamma}_{j,2}(Q_{j,i})$$

and

$$\hat{\eta}_i(t, x) = \hat{\eta}_{1,i}(t, x) - \hat{\eta}_{0,i}(t, x)$$

and  $\hat{\gamma}_{j,0}$ ,  $\hat{\gamma}_{j,1}$  and  $\hat{\gamma}_{j,2}$  are the empirical analogues of  $\gamma_{j,0}$ ,  $\gamma_{j,1}$  and  $\gamma_{j,2}$ ,  $j = \{0, 1\}$ , respectively, as defined in (15).

The proof follows two steps. In the first step in this proof is to show that

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \hat{\eta}_i(t, x) - \hat{\alpha}^{KM}(X_i; t, x) (D_i - \hat{p}_n(X_i)) \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta_i(t, x) - \alpha(X_i; t, x) (D_i - p(X_i))) + o_{\mathbb{P}}(1) \end{aligned} \quad (\text{A.21})$$

uniformly in  $(t, x) \in \mathcal{S}$ , that is, there is no estimation effect coming from replacing the true  $\eta(t, x)$ ,  $\alpha(X; t, x)$  and  $p(X)$  with their nonparametric estimators.

In the second step, we prove that, under  $H_0$ ,  $H_1$  or  $H_{1,n}$ ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta_i(t, x, p) - \alpha(X_i; t, x) (D_i - p(X_i))) V_i \quad (\text{A.22})$$

converges weakly to the same limit process as in Theorem 1.

We proceed with the proof of the first step. For  $j = \{0, 1\}$ , consider the class of measurable functions

$$\begin{aligned} \tilde{\mathcal{F}}_j = \{ & (\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}) \rightarrow \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x, p) \gamma_{j,0}(\bar{\omega}) \bar{\delta} \\ & + \gamma_{j,1}(\bar{\omega}) (1 - \bar{\delta}) - \gamma_{j,2}(\bar{\omega}) : (t, x) \in \mathcal{S}, p \in \mathcal{H}\}, \end{aligned}$$

where  $\mathcal{H}$  is the collection of all distribution functions that satisfy Assumption 7. We prove that the  $\tilde{\mathcal{F}}_j$  is Donsker. First, similar to Theorem 1, define the class of real-valued measurable functions on  $\chi_Y \times \chi_X \times \{0, 1\}$

$$\begin{aligned} \mathcal{F}_{0,3} \equiv \{ & (\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_0(\bar{\omega}, \bar{x}, \bar{z}; t, x, p) \equiv p(\bar{x}) 1\{\bar{\omega} \leq t\} \\ & \times 1\{\bar{x} \leq x\} : (t, x) \in \mathcal{S}, p \in \mathcal{H}\}, \end{aligned}$$

$$\begin{aligned} \mathcal{F}_{1,3} \equiv \{ & (\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_1(\bar{\omega}, \bar{x}, \bar{z}; t, x, p) \equiv (1 - p(\bar{x})) 1\{\bar{\omega} \leq t\} \\ & \times 1\{\bar{x} \leq x\} : (t, x) \in \mathcal{S}, p \in \mathcal{H}\}. \end{aligned}$$

Note that, for each  $((t, x), p) \in \mathcal{S} \times \mathcal{H}$ , we have that, for  $j = \{0, 1\}$ ,

$$E \left[ \sup |\xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x, p_1) - \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x, p)|^2 \right] \leq \mathcal{C} \delta^2,$$

where the supremum is over the set  $(t_1, x_1) \in \mathcal{S}$  and  $p_1 \in \mathcal{H}$  such that  $|(t_1, x_1) - (t, x)| \leq \delta$  and  $\sup_{x \in \chi_X} |p_1(x) - p(x)| \leq \delta$ , respectively. By Lemma A.1 and Theorem 19.5 in van der Vaart (1998), the classes of functions  $\mathcal{F}_{0,3}$  and  $\mathcal{F}_{1,3}$  are Donsker. Then, by Theorem 2.1 of Bae and Kim (2003), we have that  $\tilde{\mathcal{F}}_1$  and  $\tilde{\mathcal{F}}_2$  are Donsker. Therefore, by a stochastic

equicontinuity argument and the Glivenko-Cantelli Theorem

$$\sup_{(t,x) \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\eta}_i(t, x, \hat{p}_n) - \eta_i(t, x, p)) \right| = o_{\mathbb{P}}(1). \quad (\text{A.23})$$

Now, consider the class of functions

$$\begin{aligned} \mathcal{F}_{0,4} \equiv & \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \tilde{\alpha}_0(\bar{\omega}, \bar{x}, \bar{z}; t, x, p, F_0) \equiv p(\bar{x}) 1\{\bar{x} \leq x\} F_0(t|\bar{x}) (\bar{z} - p(\bar{x})) \\ & : (t, x) \in \mathcal{S}, p \in \mathcal{H}_1, F_0 \in \mathcal{H}_2\}, \end{aligned}$$

$$\begin{aligned} \mathcal{F}_{1,4} \equiv & \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \tilde{\alpha}_1(\bar{\omega}, \bar{x}, \bar{z}; t, x, p, F_1) \equiv (1 - p(\bar{x})) 1\{\bar{x} \leq x\} F_1(t|\bar{x}) (\bar{z} - p(\bar{x})) \\ & : (t, x) \in \mathcal{S}, p \in \mathcal{H}_1, F_q \in \mathcal{H}_2\}. \end{aligned}$$

Again, for each  $((t, x), p, F_j) \in \mathcal{S} \times \mathcal{H}_1 \times \mathcal{H}_2$ , we have that, for  $j = \{0, 1\}$ ,

$$E \left[ \sup |\tilde{\alpha}_j(\bar{\omega}, \bar{x}, \bar{z}; t_1, x_1, p_1; F_{j,1}) - \tilde{\alpha}_j(\bar{\omega}, \bar{x}, \bar{z}; t, x, p; F_j)|^2 \right] \leq \mathcal{C}\delta^2,$$

where the supremum is over the set  $(t_1, x_1) \in \mathcal{S}$ ,  $p_1 \in \mathcal{H}_1$  and  $F_j \in \mathcal{H}_2$  such that  $|(t_1, x_1) - (t, x)| \leq \delta$ ,  $\sup_{x \in \mathcal{X}_X} |p_1(x) - p(x)| \leq \delta$  and  $\sup_{(t,x) \in \mathcal{S}} |F_{j,1}(t|x) - F_j(t|x)| \leq \delta$  respectively. By Lemma A.1 and Theorem 19.5 in van der Vaart (1998), the classes of functions  $\mathcal{F}_{0,4}$  and  $\mathcal{F}_{1,4}$  are Donsker. Therefore, by a stochastic equicontinuity argument, the Glivenko-Cantelli Theorem and the triangle inequality, we have

$$\sup_{(t,x) \in \mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \hat{\alpha}^{KM}(X_i; t, x) (D_i - \hat{p}(X_i)) - \alpha(X_i; t, x) (D_i - p(X_i)) \right) \right| = o_{\mathbb{P}}(1). \quad (\text{A.24})$$

Combining (A.23) and (A.24), we have established (A.21), finishing the proof of the first step.

Next, let's consider (A.22). Define the classes of real measurable functions

$$\begin{aligned} \mathcal{G}_{0,1,*} \equiv & \{(\bar{w}, \bar{x}, \bar{z}, \bar{\delta}, \bar{v}) \in \mathcal{X}_Y \times \mathcal{X}_X \times \{0, 1\} \times \{0, 1\} \times \mathcal{X}_v \rightarrow g_0(\bar{w}, \bar{x}, \bar{z}, \bar{\delta}, \bar{v}; t, x) \equiv \\ & (1\{\bar{w} \leq t\} 1\{\bar{x} \leq x\} p(\bar{x}) \gamma_{0,0}(\bar{w}) \bar{\delta} + \gamma_{0,1}(\bar{w}) (1 - \bar{\delta}) - \gamma_{0,2}(\bar{w}) \\ & + (1 - p(\bar{x})) 1\{\bar{x} \leq x\} F_0(y|X_i) (\bar{z} - p(\bar{x}))) \bar{v} : (t, x) \in \mathcal{S}\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{G}_{1,1,*} \equiv & \{(\bar{w}, \bar{x}, \bar{z}, \bar{\delta}, \bar{v}) \in \mathcal{X}_Y \times \mathcal{X}_X \times \{0, 1\} \times \{0, 1\} \times \mathcal{X}_v \rightarrow g_1(\bar{w}, \bar{x}, \bar{z}, \bar{\delta}, \bar{v}; t, x) \equiv \\ & (1\{\bar{w} \leq t\} 1\{\bar{x} \leq x\} (\bar{z} - p(\bar{x})) \gamma_{1,0}(\bar{w}) \bar{\delta} + \gamma_{1,1}(\bar{w}) (1 - \bar{\delta}) - \gamma_{1,2}(\bar{w}) \\ & - p(\bar{x}) 1\{\bar{x} \leq x\} F_1(y|\bar{x}) (\bar{z} - p(\bar{x}))) \bar{v} : (t, x) \in \mathcal{S}\}. \end{aligned}$$

For  $j = \{0, 1\}$ , the classes  $\mathcal{G}_{j,1,*}$  are  $P_{(\bar{w}, \bar{x}, \bar{z}, \bar{\delta}, \bar{v})}$ -Donsker, since  $\mathcal{G}_{j,1}$  are  $P_{\bar{w}, \bar{x}, \bar{z}, \bar{\delta}}$ -Donsker, see Theorem 2.9.6 in van der Vaart and Wellner (1996). Then, since  $\mathbb{P}_n^* g_j = 0$  for all  $g_j \in \mathcal{G}_{j,1,*}$ ,

$$I^*(t, x) = \frac{1}{n} \sum_{i=1}^n (\eta_i(t, x) - \alpha(X_i; t, x) (D_i - p(X_i))) V_i + o_{\mathbb{P}_n^*}(n^{-1/2}), \quad (\text{A.25})$$

uniformly in  $(t, x) \in \mathcal{S}$ .

The expansion (A.25), and the multiplier functional central limit theorem, see Theorem 2.9.6 in van der Vaart and Wellner (1996), imply that  $\sqrt{n}I^*(t, x)$  converges weakly (almost surely) to the same weak limit as  $\sqrt{n}\hat{I}(t, x)$  in  $l^\infty(\mathcal{S})$  under  $H_0$ ,  $H_1$  or  $H_{1n}$ .

This completes the proof of Theorem 4. ■

**Proof of Theorem 5** First, we must derive the asymptotic linear representation of the process  $(\hat{I}_\tau^{CATE} - I_\tau^{CATE})(x)$ . To consider the most general case, we set  $\tau = \tau_C$ . Then, we can rewrite  $\hat{I}_\tau^{CATE}(\cdot)$  as

$$\begin{aligned} \hat{I}_\tau^{CATE} &= \bar{I}_\tau^{CATE}(x) \\ &+ \frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1^{CATE}(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t, x) - \xi_1^{CATE}(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; x) \right) \\ &- \frac{n_0}{n} \sum_{l=1}^{n_0} W_{0,l:n_0} \left( \hat{\xi}_0^{CATE}(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x) - \xi_0^{CATE}(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; x) \right), \end{aligned} \quad (\text{A.26})$$

where  $\bar{I}_\tau^{CATE}(x)$  is defined similarly to (A.1) but replacing  $\xi_1$  and  $\xi_1$  with

$$\begin{aligned} \xi_1^{CATE}(\bar{y}, \bar{x}, \bar{z}; x) &= \bar{z}(1 - p(\bar{x}))\bar{y}1\{\bar{x} \leq x\}, \\ \xi_0^{CATE}(\bar{y}, \bar{x}, \bar{z}; x) &= (1 - \bar{z})p(\bar{x})\bar{y}1\{\bar{x} \leq x\}. \end{aligned}$$

Additionally,  $\hat{\xi}_1^{CATE}$  and  $\hat{\xi}_0^{CATE}$  are defined similarly to  $\xi_1^{CATE}$  and  $\xi_0^{CATE}$ , but replacing the true propensity score  $p(\cdot)$  by the SLE  $\hat{p}(\cdot)$ .

We will derive the uniform representation of each term separately, as in Theorem 1. To this end, define the classes of real-value measurable functions on  $\chi_Y \times \chi_X \times \{0, 1\}$

$$\mathcal{H}_{0,1} \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_0^{CATE}(\bar{y}, \bar{x}, \bar{z}; x) \equiv p(\bar{x})\bar{\omega}1\{\bar{x} \leq x\} : x \in \mathbb{R}^k\},$$

$$\mathcal{H}_{1,1} \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \rightarrow \xi_1^{CATE}(\bar{y}, \bar{x}, \bar{z}; x) \equiv (1 - p(\bar{x}))\bar{\omega} \times 1\{\bar{x} \leq x\} : x \in \mathbb{R}^k\}$$

Notice that  $\mathcal{H}_{j,1}$  are a VC-subgraph classes of functions with VC index smaller or equal than  $k + 2$  and admits the envelope  $\Phi(\bar{\omega}, \bar{x}, \bar{z}) = |\bar{\omega}|$  that satisfies, under Assumption 10, the required moment conditions of Theorem 1 of Sellero et al. (2005). Thus,

$$\bar{I}_\tau^{CATE}(x) = \frac{1}{n} \sum_{i=1}^n \eta_i^{CATE}(x) + R_n^{CATE}(x) \quad (\text{A.27})$$

where

$$\eta_i^{CATE}(x) = \eta_{1,i}^{CATE}(x) - \eta_{0,i}^{CATE}(x),$$

and for  $j = \{0, 1\}$ ,

$$\eta_{j,i}^{CATE}(x) = \xi_j^{CATE}(Q_i, X_i, D_i; x) \gamma_{j,0}(Q_{j,i}) \delta_{j,i} + \gamma_{j,1}^{CATE}(Q_{j,i})(1 - \delta_{j,i}) - \gamma_{j,2}^{CATE}(Q_{j,i}),$$

$$\begin{aligned}
\gamma_{j,0}(\bar{t}) &= \exp \left\{ \int_0^{\bar{t}^-} \frac{H_{j,0}(d\bar{w})}{1 - H_j(\bar{w})} \right\}, \\
\gamma_{j,1}(\bar{t}) &= \frac{1}{1 - H_j(\bar{t})} \int 1 \{ \bar{t} < \bar{w} \} \xi_j^{CATE}(\bar{w}, \bar{x}, \bar{z}; x) \gamma_{j,0}(\bar{w}) H_{j,11}(d\bar{w}, d\bar{x}), \\
\gamma_{j,2}(\bar{t}) &= \int \int \frac{1 \{ \bar{v} < \bar{t}, \bar{v} < \bar{w} \} \xi_j^{CATE}(\bar{w}, \bar{x}, \bar{z}; x)}{[1 - H_j(\bar{v})]^2} \gamma_{j,0}(\bar{w}) H_{j,0}(d\bar{v}) H_{j,11}(d\bar{w}, d\bar{x}).
\end{aligned}$$

and

$$\sup_{x \in \mathbb{R}^k} |R_n^{CATE}(x)| = O\left(\frac{\ln^3 n}{n}\right) \text{ a.s.}$$

Now, we look for the second term of (A.26). Using similar arguments as in the proof of Lemma 1, we can establish that

$$\begin{aligned}
& \frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left( \hat{\xi}_1^{CATE}(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t, x) - \xi_1^{CATE}(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; x) \right) \\
& - \frac{n_0}{n} \sum_{l=1}^{n_0} W_{0,l:n_0} \left( \hat{\xi}_0^{CATE}(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x) - \xi_0^{CATE}(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; x) \right) \\
& = \frac{1}{n} \sum_{i=1}^n \alpha^{CATE}(X_i; x) (D_i - p(X_i)) + o_{\mathbb{P}}(n^{-1/2}) \tag{A.28}
\end{aligned}$$

uniformly in  $x \in \mathcal{W}_X$ , where

$$\alpha^{CATE}(\bar{x}; x) = \alpha_1^{CATE}(\bar{x}; x) - \alpha_0^{CATE}(\bar{x}; x)$$

and

$$\begin{aligned}
\alpha_1^{CATE}(\bar{x}; x) &= -p(\bar{x}) 1 \{ \bar{x} \leq x \} \mathbb{E}(Y_1 | \bar{x}), \\
\alpha_0^{CATE}(\bar{x}; x) &= (1 - p(\bar{x})) 1 \{ \bar{x} \leq x \} \mathbb{E}(Y_0 | \bar{x}).
\end{aligned}$$

Combining (A.27) and (A.28), we conclude that

$$\begin{aligned}
\hat{I}_\tau^{CATE} - I_\tau^{CATE} &= \frac{1}{n} \sum_{i=1}^n \left[ \left( \eta_i^{CATE}(x) - I_\tau^{CATE}(x) \right) \right. \\
& \quad \left. - \alpha^{CATE}(X_i; x) (D_i - p(X_i)) \right] + o_{\mathbb{P}}(n^{-1/2}) \tag{A.29}
\end{aligned}$$

uniformly in  $x \in \mathcal{W}_X$ , concluding the proof of the asymptotic linear representation.

Once we have proved the validity of the uniform linear representation (A.29), the proof of the weak converge of the process  $\sqrt{n}(\bar{I}_n^\tau - I^\tau)(x)$  under  $H_0^{CATE}$ ,  $H_1^{CATE}$  and  $H_{1,n}^{CATE}$  follows the same steps of Theorems 1, 2 and 3, and the validity of the bootstrap follows the reasoning of Theorem 4 in a routine fashion. Details are omitted. ■

## References

- Abadie, A. (2002), “Bootstrap tests for distributional treatment effects in instrumental variable models,” *Journal of the American Statistical Association*, 97(457), 284–292.
- Abbring, J. H., and van den Berg, G. J. (2003), “The nonparametric identification of treatment effects in duration models,” *Econometrica*, 71(5), 1491–1517.
- Abbring, J. H., and van den Berg, G. J. (2005), “Social experiments and instrumental variables with duration outcomes,” *IFS Working Papers*, 19.
- Akritis, M. G. (1994), “Nearest neighbor estimation of a bivariate distribution under random censoring,” *The Annals of Statistics*, 22(3), 1299–1327.
- Andrews, D. W. K., and Shi, X. (2013), “Inference Based on Conditional Moment Inequalities,” *Econometrica*, 81(2), 609–666.
- Andrews, D. W. K., and Shi, X. (2014), “Nonparametric inference based on conditional moment inequalities,” *Journal of Econometrics*, 179(1), 31–45.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), “Identification of causal effects using instrumental variables,” *Journal of the American Statistical Association*, 91(434), 444–455.
- Bae, J., and Kim, S. (2003), “The uniform central limit theorem for the Kaplan-Meier integral process,” *Bulletin of the Australian Mathematical Society*, 67, 467–480.
- Barrett, G., and Donald, S. G. (2003), “Consistent tests for stochastic dominance,” *Econometrica*, 71(1), 71–104.
- Bierens, H. J., and Ploberger, W. (1997), “Asymptotic theory of integrated conditional moment tests,” *Econometrica*, 65(5), 1129–1151.
- Bijwaard, G. E., and Ridder, G. (2005), “Correcting for selective compliance in a re-employment bonus experiment,” *Journal of Econometrics*, 125, 77–111.
- Bitler, M. O., Gelbach, J. B., and Hoynes, H. H. (2006), “What mean impacts miss: Distributional effects of welfare reform experiments,” *The American Economic Review*, 96(4), 988–1012.
- Bitler, M. O., Gelbach, J. B., and Hoynes, H. H. (2014), Can Variation in Subgroups’ Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment,.
- Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2008), “Distributional impacts of the Self-Sufficiency Project,” *Journal of Public Economics*, 92(3-4), 748–765.
- Breunig, C., Mammen, E., and Simoni, A. (2014), “Semi/Nonparametric Estimation in case of Endogeneous Selection,” *Mimeo*, .
- Cabus, P. (1998), “Un test de type Kolmogorov-Smirnov dans le cadre de comparaison de fonctions de régression,” *Comptes Rendus de l’Académie des Sciences, Series I*(327), 939–942.

- Chen, K. (2001), “Parametric models for response-biased sampling,” *Journal of the Royal Statistical Society: Series B*, 63(4), 775–789.
- Chen, X., Linton, O., and Van Keilegom, I. (2003), “Estimation of semiparametric models when the criterion function is not smooth,” *Econometrica*, 71(5), 1591–1608.
- Cox, D. R. (1972), “Regression models and life-tables (with discussion),” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2), 187–220.
- Crépon, B., Ferracci, M., Jolivet, G., and van den Berg, G. J. (2009), “Active labor market policy effects in a dynamic setting,” *Journal of the European Economic Association*, 7(May 2009), 595–605.
- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008), “Nonparametric tests for treatment effect heterogeneity,” *The Review of Economics and Statistics*, 90(3), 389–405.
- Delgado, M. A. (1993), “Testing the equality of nonparametric regression curves,” *Statistics & Probability Letters*, 17(3), 199–204.
- Delgado, M. a., and Escanciano, J. C. (2013), “Conditional stochastic dominance testing,” *Journal of Business & Economic Statistics*, 31(1), 16–28.
- Delgado, M. A., and González-Manteiga, W. (2001), “Significance testing in nonparametric regression based on the bootstrap,” *The Annals of Statistics*, 29(5), 1469–1507.
- D’Haultfoeuille, X. (2010), “A new instrumental method for dealing with endogenous selection,” *Journal of Econometrics*, 154(1), 1–15.
- Donald, S. G., and Hsu, Y.-C. (2013), “Estimation and inference for distribution functions and quantile functions in treatment effect models,” *Journal of Econometrics*, .
- Eberwein, C., Ham, J. C., and Lalonde, R. J. (1997), “The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data,” *The Review of Economic Studies*, 64(4), 655–682.
- Escanciano, J. C. (2006a), “A consistent diagnostic test for regression models using projections,” *Econometric Theory*, 22, 1030–1051.
- Escanciano, J. C. (2006b), “Goodness-of-Fit Tests for Linear and Nonlinear Time Series Models,” *Journal of the American Statistical Association*, 101(474), 531–541.
- Frandsen, B. R. (2014), “Treatment Effects With Censoring and Endogeneity,” *Mimeo*, .
- Gill, R. (1983), “Large sample behaviour of the product-limit estimator on the whole line,” *The Annals of Statistics*, 11(1), 49–58.
- Giné, E., and Zinn, J. (1990), “Bootstrapping general empirical measures,” *The Annals of Probability*, 18(2), 851–869.
- González-Manteiga, W., and Cadarso-Suárez, C. (1994), “Asymptotic properties of a generalized kaplan-meier estimator with some applications,” *Journal of Nonparametric Statistics*, 4(1), 65–78.

- Grambsch, P., and Therneau, T. (1994), “Proportional hazards tests and diagnostics based on weighted residuals,” *Biometrika*, 81(3), 515–526.
- Hahn, J. (1998), “On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,” *Econometrica*, 66(2), 315–331.
- Ham, J. C., and Lalonde, R. J. (1996), “The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training,” *Econometrica*, 64(1), 175–205.
- Heckman, J. J., and Vytlacil, E. J. (2007), “Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation,” *Handbook of Econometrics*, 6B(70), 4779–4874.
- Hirano, K., Imbens, G. W., and Ridder, G. (2003), “Efficient estimation of average treatment effects using the estimated propensity score,” *Econometrica*, 71(4), 1161–1189.
- Honore, B., Khan, S., and Powell, J. (2002), “Quantile regression under random censoring,” *Journal of Econometrics*, 109, 67–105.
- Hsu, Y.-C. (2013), “Consistent tests for conditional treatment effects,” *Mimeo*, .
- Hubbard, A. E., Laan, M. J. V. D., and Robins, J. M. (2000), “Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies,” in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, eds. M. Halloran, and D. Berry, Vol. 116, New York: Springer, pp. 135–177.
- Ichimura, H., and Linton, O. (2005), “Asymptotic expansions for some semiparametric program evaluation estimators,” in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews, and J. H. Stock Cambridge University Press, chapter 8, pp. 149–170.
- Iglesias Pérez, C., and González-Manteiga, W. (1999), “Strong representation of a generalized product-limit estimator for truncated and censored data with some applications,” *Journal of Nonparametric Statistics*, 10(3), 213–244.
- Imbens, G. W. (2004), “Nonparametric estimation of average treatment effects under exogeneity: A review,” *The Review of Economics and Statistics*, 86(1), 4–29.
- Imbens, G. W., and Angrist, J. D. (1994), “Identification and estimation of local average treatment effects,” *Econometrica*, 62(2), 467–475.
- Imbens, G. W., and Wooldridge, J. M. (2009), “Recent developments in the econometrics of program evaluation,” *Journal of Economic Literature*, 47(1), 5–86.
- Kaplan, E. L., and Meier, P. (1958), “Nonparametric estimation from incomplete observations,” *Journal of the American Statistical Association*, 53(282), 457–481.
- Khan, S., and Tamer, E. (2010), “Irregular Identification, Support Conditions, and Inverse Weight Estimation,” *Econometrica*, 78(6), 2021–2042.

- Koul, H. L., and Stute, W. (1999), “Nonparametric model checks for time series,” *The Annals of Statistics*, 27(1), 204–236.
- Lee, M.-j. (2009), “Non-parametric tests for distributional treatment effect for randomly censored responses,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1), 243–264.
- Lee, M.-j., and Lee, S.-j. (2005), “Analysis of job-training effects on Korean women,” *Journal of Applied Econometrics*, 20(4), 549–562.
- Lee, S., Song, K., and Whang, Y.-J. (2013), “Testing functional inequalities,” *Journal of Econometrics*, 172(1), 14–32.
- Lee, S., and Whang, Y. (2009), “Nonparametric tests of conditional treatment effects,” *Cowles Foundation Discussion Paper*, 1740.
- Li, Q., Racine, J. S., and Wooldridge, J. M. (2009), “Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data,” *Journal of Business & Economic Statistics*, 27(2), 206–223.
- Mammen, E. (1993), “Bootstrap and wild bootstrap for high dimensional linear models,” *The Annals of Statistics*, 21(1), 255–285.
- Meyer, B. D. (1996), “What have we learned from the Illinois reemployment bonus experiment?,” *Journal of Labor Economics*, 14(1), 26–51.
- Neumeyer, N., and Dette, H. (2003), “Nonparametric comparison of regression curves: an empirical process approach,” *The Annals of Statistics*, 31(3), 880–920.
- Newey, W. K. (1997), “Convergence rates and asymptotic normality for series estimators,” *Journal of Econometrics*, 79(1), 147–168.
- Pardo-Fernandez, J. C., and Van Keilegom, I. (2006), “Comparison of regression curves with censored responses,” *Scandinavian Journal of Statistics*, 33(3), 409–434.
- Rosenbaum, P. R., and Rubin, D. B. (1983), “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies,” *Journal of Educational Psychology*, 66(5), 688–70.
- Satten, G. A., and Datta, S. (2001), “The KaplanMeier estimator as an inverse-probability-of-censoring weighted average,” *The American Statistician*, 55(3), 207–210.
- Sellero, C. S., González-Manteiga, W., and van Keilegom, I. (2005), “Uniform representation of product-limit integrals with applications,” *Scandinavian Journal of Statistics*, 32, 563–581.
- Sianesi, B. (2004), “An evaluation of the Swedish system of active labor market programs in the 1990s,” *The Review of Economics and Statistics*, 86(1), 133–155.
- Srihera, R., and Stute, W. (2010), “Nonparametric comparison of regression functions,” *Journal of Multivariate Analysis*, 101(9), 2039–2059.

- Stinchcombe, M. B., and White, H. (1998), “Consistent specification testing with nuisance parameters present only under the alternative,” *Econometric theory*, 14, 295–325.
- Stute, W. (1993), “Consistent estimation under random censorship when covariables are present,” *Journal of Multivariate Analysis*, 45, 89 – 103.
- Stute, W. (1995), “The central limit theorem under random censorship,” *The Annals of Statistics*, 23(2), 422–439.
- Stute, W. (1996), “Distributional convergence under random censorship when covariables are present,” *Scandinavian Journal of Statistics*, 23(4), 461–471.
- Stute, W. (1997), “Nonparametric model checks for regression,” *The Annals of Statistics*, 25(2), 613–641.
- Stute, W., González-Manteiga, W., and Quindimil, M. P. (1998), “Bootstrap Approximations in Model Checks for Regression,” *Journal of the American Statistical Association*, 93(441), 141–149.
- Stute, W., Manteiga, W. G., and Sellero, C. S. (2000), “Nonparametric model checks in censored regression,” *Communications in Statistics - Theory and Methods*, 29(7), 1611 – 1629.
- Stute, W., and Wang, J.-L. (1993a), “Multi-sample U-statistics for censored data,” *Scandinavian journal of statistics*, 20(4), 369–374.
- Stute, W., and Wang, J.-L. (1993b), “The strong law under random censorship,” *The Annals of Statistics*, 21(3), 1591–1607.
- Tang, G., Little, R., and Raghunathan, T. (2003), “Analysis of multivariate missing data with nonignorable nonresponse,” *Biometrika*, 90(4), 747–764.
- van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.
- Woodbury, S. A., and Spiegelman, R. G. (1987), “Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois,” *The American Economic Review*, 77(4), 513–530.
- Zhou, M. (1992), “Asymptotic Normality of the ‘Synthetic Data’ Regression Estimator for Censored Survival Data,” *The Annals of Statistics*, 20(2), 1002–1021.