PROGRAM EVALUATION WITH RIGHT-CENSORED DATA: SUPPLEMENTAL APPENDIX

Pedro H. C. Sant'Anna

This appendix contains material to supplement the paper "Program Evaluation with Right-Censored Data", by Pedro H. C. Sant'Anna. Section A discuss how one can construct a test for the assumptions about the censoring mechanism. Section B presents detailed discussion on how we construct the estimators used in the Monte Carlo in Section 6 of the paper. Section C collects the proofs of the general results presented in Sections 3 and 4 of the paper. Section D collects the proofs of the policy evaluation parameters discussed in Section 5 of the paper. More precisely, Sections D.1, D.2 and D.3 present the proofs of Propositions 2, 3 and 4, respectively. Finally, auxiliary lemmas are collected in Section E.

A. Model Specification Test

Notice that all our asymptotic results on the 2SKM estimator rely on covariates not providing any additional information if censoring will take place or not, that is, Theorems 1, 2 and 3 depend on the validity of Assumption 3.1. As pointed out in Remark 2, such an assumption can be relaxed if the data can be partitioned into groups/clusters. However, in some situations, such a task is not feasible or is hard to be theoretically justified. Thus, in order to assess the "reliability" of results based on our 2SKM procedure, it would be desirable to test the validity of Assumption 3.1 as a form of specification test. Given that our 2SKM estimator is suitable under Assumption 3.1 but may be questionable under the weaker condition that $Y \perp\!\!\!\perp C|X,T$, in this section we briefly discuss how one can nonparametrically test the former against the later. That is, we wish to test

$$H_0 : Y \perp\!\!\!\perp C|T \textit{ and } \mathbb{P}\left(\delta = 1|X,Y,T\right) = \mathbb{P}\left(\delta = 1|Y,T\right) \ a.s.,$$

Date: First version: November 2015. This version: April 7, 2016.

Department of Economics, Vanderbilt University, 415 Calhoun Hall, Nashville, TN, USA. Email: pedro.h.santanna@vanderbilt.edu. I gratefully acknowledge research support from Spanish Plan Nacional de I+D+I, reference number ECO2014-55858-P.

against

$$H_1 : Y \perp\!\!\!\perp C | X, T.$$

Notice that we are interested in testing Assumption 3.1 against a particular direction, $Y \perp\!\!\!\perp C|X, T$. The reason for this is that the (joint) distribution of (Y, X, T) is not nonparametrically (point) identified when H_1 is false and no additional assumptions are imposed, cf. Tsiatis (1975) and Heckman and Honoré (1989). Thus, one may argue that without relying on external data or additional restrictions, testing H_0 against H_1 is the "best one can hope for".

A simple way of testing H_0 against H_1 is to compare estimates of $F(y, x, t)$ under H_0 and H_1. We know from Proposition 1 and Corollary 1 that, under H_0, $F(y, x, t)$ can be estimated by $\hat{F}_n^{km}(y, x, t)$ as defined in (3.4). From Dabrowska (1989), González-Manteiga and Cadarso-Suárez (1994) and Lopez (2011), we have that, under H_1 and additional smoothness conditions, $F(y, x, t)$ can be nonparametrically estimated by

$$\tilde{F}_n(y, x, t) = \sum_{j=0}^{t} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i 1\{Q_i \le y\} 1\{X_i \le x\} 1\{T_i = j\}}{1 - G_n(Q_i - |X_i, T_i = j)},$$

where $G_n(y|x, t)$ is the Beran (1981)-type estimator for $\mathbb{P}(C \le y|X = x, T = t)$,

$$G_n(y|x, t) = 1 - \prod_{Q_i \le y, T_i = t, \delta_i = 0}^{n} \left(1 - \frac{\omega_{in}(x)}{\sum_{j=1}^{n} \omega_{jn}(x) 1\{T_j = t\} 1\{Q_j \ge Q_i\}}\right),$$

and $\omega_{in}(x)$ are the Nadaraya–Watson weights,

$$\omega_{in}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j - x}{h}\right)},$$

with K being a (high order) positive kernel function and h a bandwidth.

Thus, one can construct different test statistics for H_0 against H_1 by comparing how close $\sqrt{n}\left(\hat{F}_n^{km}(y, x, t) - \tilde{F}_n(y, x, t)\right)$ is to zero. For instance, one can consider the Kolmogorov-Smirnov (KS) test statistic

$$KS_n = \sqrt{n} \sup_{(y,x,t) \in \mathcal{Z}} \left| \hat{F}_n^{km}(y, x, t) - \tilde{F}_n(y, x, t) \right|,$$

where $\mathcal{Z} \in (-\infty, \tau_H) \times \mathbb{R}^k \times \{0, 1\}$[10]. Alternatively, one can use the Cramér-von Mises, or projection based tests, cf. Stute (1997), Stute et al. (1998), and Escanciano (2006).

On one hand, if the test statistic is smaller than the (potentially bootstrapped) critical value, one would not reject H_0, suggesting that the proposed 2SKM procedure is adequate. On the other hand, if the test statistic is sufficiently large, it would indicate that one should be cautious in using 2SKM estimators.

Here, it is important to emphasize that in order to establish the asymptotic validity of likelihood ratio (LR)-type tests such as KS_n, one must impose additional smoothness assumptions on the DGP due to the fact that $\tilde{F}_n(y, x, t)$ is based on smoothed estimates, cf. Dabrowska (1989), González-Manteiga and Cadarso-Suárez (1994), and Lopez (2011). Given that such assumptions are not necessary to derive the asymptotic properties of $\left(\hat{F}_n^{km} - F\right)(\cdot, \cdot, \cdot)$, one must be cautious when the test statistic is sufficiently large due to the fact that this would indicate rejection of Assumption 3.1 only if such additional smoothness assumptions are satisfied[11].

B. Details about the Monte Carlo

In this section, we provide further details on the Monte Carlo exercises employed in our simulation study, by describing how we construct the policy evaluation parameters using the different competing methods, and presenting further numerical results not presented in the paper.

B.1. Competing methods

In the Monte Carlo, we compare our 2SKM proposal with three other competing methods: one that ignores that the outcome of interested is subjected to censoring (we label such an approach as "Ignore "); one that is based only on uncensored data (we label such an approach as "Uncens "); one based on the Cox (1972, 1975) Proportional hazard model (we label such an approach as "Cox"); and Frandsen

10 We note that Frandsen (2015a) propose a test statistic similar to KS_n, but in a context without covariates and when the censoring points C are always observed. The availability of covariates allows one to relax the requirement that C is available even for the uncesored individuals.

11 A detailed derivation of the asymptotic properties of LR-type tests such as KS_n is beyond the scope of this paper and is deferred to future work.

(2015b)'s proposal (we label such an approach as "Frandsen"). The comparisons are based on the following measures of interest: $\mathbb{E}(Y_1)$, $\mathbb{E}(Y_0)$, $F_{Y_1}^{-1}(0.5)$, $F_{Y_0}^{-1}(0.5)$, ATE and $QTE(0.5)$. Estimation of these measures using the 2SKM has already been discussed in Section 5.1. In the following, we describe how we estimate these measures using the "Ignore", "Uncens", "Cox" and "Frandsen" approaches.

For the "Ignore" approach, one simply ignores that the outcome of interest is censored, and uses the inverse probability weighted (IPW) estimators for $\mathbb{E}(Y_1)$, $\mathbb{E}(Y_0)$, and ATE proposed by Hirano et al. (2003), and those for $F_{Y_1}^{-1}(0.5)$, $F_{Y_0}^{-1}(0.5)$, and $QTE(0.5)$ proposed by Donald and Hsu (2014). That is,

$$
\begin{aligned}
\mathbb{E}_n^{ignore}[Y_0] &= \frac{1}{n}\sum_{i=1}^{n}\frac{(1-T_i)\,Q_i}{1-\hat{p}_n(X_i)},\\
\mathbb{E}_n^{ignore}[Y_1] &= \frac{1}{n}\sum_{i=1}^{n}\frac{T_iQ_i}{\hat{p}_n(X_i)},\\
ATE_n^{ignore} &= \mathbb{E}_n^{ignore}[Y_1]-\mathbb{E}_n^{ignore}[Y_0]
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{F}_{n,Y_0}^{ignor,-1}(0.5) &= \inf\left\{y:\hat{F}_{n,Y_0}^{ignore,r}(y)\geq 0.5\right\},\\
\hat{F}_{n,Y_1}^{ignor,-1}(0.5) &= \inf\left\{y:\hat{F}_{n,Y_1}^{ignore,r}(y)\geq 0.5\right\},\\
QTE_n^{ignore} &= \hat{F}_{n,Y_1}^{ignor,-1}(0.5)-\hat{F}_{n,Y_0}^{ignor,-1}(0.5),
\end{aligned}
$$

where $\hat{F}_{n,Y_0}^{ignore,r}(y)$ and $\hat{F}_{n,Y_1}^{ignore,r}(y)$ denotes the rearrangement of

$$
\begin{aligned}
\hat{F}_{n,Y_0}^{ignore}(y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{(1-T_i)\,1\{Q_i\leq y\}}{1-\hat{p}_n(X_i)},\\
\hat{F}_{n,Y_1}^{ignore}(y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{T_i1\{Q_i\leq y\}}{\hat{p}_n(X_i)},
\end{aligned}
$$

respectively, if $\hat{F}_{n,Y_0}^{ignore}(y)$ and $\hat{F}_{n,Y_1}^{ignore}(y)$ are not monotone, cf. Chernozhukov et al. (2010), and $\hat{p}_n(\cdot)$ is the Logit Series estimator for the propensity score $p(\cdot)$.

The "Uncens" approach follows the same idea of the "Ignore" approach, but using only the uncensored data, leading to the following estimators:

$$\begin{aligned}
\mathbb{E}_n^{uncens}[Y_0] &= \frac{1}{n_{uncens}} \sum_{i=1}^{n} \delta_i \frac{(1-T_i) Q_i}{1-\hat{p}_n(X_i)}, \\
\mathbb{E}_n^{uncens}[Y_1] &= \frac{1}{n_{uncens}} \sum_{i=1}^{n} \delta_i \frac{T_i Q_i}{\hat{p}_n(X_i)}, \\
ATE_n^{uncens} &= \mathbb{E}_n^{uncens}[Y_1] - \mathbb{E}_n^{uncens}[Y_0]
\end{aligned}$$

and

$$\begin{aligned}
\hat{F}_{n,Y_0}^{uncens,-1}(0.5) &= \inf\left\{ y : \hat{F}_{n,Y_0}^{uncens,r}(y) \geq 0.5 \right\}, \\
\hat{F}_{n,Y_1}^{uncens,-1}(0.5) &= \inf\left\{ y : \hat{F}_{n,Y_1}^{uncens,r}(y) \geq 0.5 \right\}, \\
QTE_n^{uncens} &= \hat{F}_{n,Y_1}^{uncens,-1}(0.5) - \hat{F}_{n,Y_0}^{uncens,-1}(0.5),
\end{aligned}$$

where $n_{uncens} = \sum_{i=1}^{n} \delta_i$, and $\hat{F}_{n,Y_0}^{uncens,r}(y)$ and $\hat{F}_{n,Y_1}^{uncens,r}(y)$ denotes the rearrangement of

$$\begin{aligned}
\hat{F}_{n,Y_0}^{uncens}(y) &= \frac{1}{n_{uncens}} \sum_{i=1}^{n} \delta_i \frac{(1-T_i) 1\{Q_i \leq y\}}{1-\hat{p}_n(X_i)}, \\
\hat{F}_{n,Y_1}^{uncens}(y) &= \frac{1}{n_{uncens}} \sum_{i=1}^{n} \delta_i \frac{T_i 1\{Q_i \leq y\}}{\hat{p}_n(X_i)},
\end{aligned}$$

respectively, if $\hat{F}_{n,Y_0}^{uncens}(y)$ and $\hat{F}_{n,Y_1}^{uncens}(y)$ are not monotone.

The "Cox" approach relies on recovering the unconditional CDF's of Y_0 and Y_1 from the conditional hazard rate of the treated and control subpopulations. Such a procedure depends on several steps. Denote by $h_0(y|x)$ and $h_1(y|X=x)$ the conditional hazard rates for the control and treated groups, respectively. The first step is to assume that $h_t(y|X=x)$, $t \in \{0,1\}$, follows the Cox model specification

$$h_t(y|X=x) = h_t^0(y) e^{x'\beta_t}, \tag{B.1}$$

where, $h_t^0(y), t \in \{0,1\}$ is the baseline hazard function for the $\{T=t\}$ subpopulation. Then we estimate β_t by $\hat{\beta}_{n,t}$, its partial-likelihood estimator, cf. Cox (1975), and the baseline cumulative hazard, $H_t(y) = \int_0^y h_t^0(\bar{y})\, d\bar{y}$, by the Breslow (1972) estimator

$$\hat{H}_{n,t}(y) = \sum_{i=1}^{n} \frac{1\{Q_i \leq y\}\delta_i}{\sum_{j=1}^{n} 1\{Q_j \geq Q_i\} \exp\left(X_{j,t}'\hat{\beta}_{n,t}\right)}.$$

Next, we exploit that, for $t \in \{0,1\}$,

$$\begin{aligned} F_{Y_t}^{cox}(y|X=x) &= 1 - \exp\left(-\int_{-\infty}^{y} h_t(\bar{y}|X=x)\, d\bar{y}\right) \\ &= 1 - \exp\left(-H_t(y)\exp^{x'\beta_t}\right), \end{aligned} \tag{B.2}$$

where the first equality follows from (Shorack and Wellner, 1986, Proposition 1, pg. 301), and the second one follows from (B.1), cf. Portnoy (2003). Thus, by plugging-in $\hat{\beta}_{n,t}$ and $\hat{H}_{n,t}(y)$ into (B.2), we estimate the conditional CDF of Y_t, $t \in \{0,1\}$, by

$$\hat{F}_{n,Y_t}^{cox}(y|X=x) = 1 - \exp\left(-\hat{H}_{n,t}(y)\exp^{x'\hat{\beta}_{n,t}}\right).$$

Given that, for $t \in \{0,1\}$,

$$F_{Y_t}^{cox}(y) = \int F_{Y_t}^{cox}(y|X=x)\, F_X(dx),$$

in the third step we simply integrate out the covariate vector, and estimate the unconditional CDF of Y_t, $t \in \{0,1\}$, by

$$\hat{F}_{n,Y_t}^{cox}(y) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{n,Y_t}^{cox}(y|X=X_i).$$

With $\hat{F}_{n,Y_t}^{cox}(y)$ at hands, we can finally estimate our measures of interest by

$$\begin{aligned} \mathbb{E}_n^{cox}(Y_0) &= \int y \hat{F}_{n,Y_0}^{cox}(dy), \\ \mathbb{E}_n^{cox}(Y_1) &= \int y \hat{F}_{n,Y_1}^{cox}(dy), \\ ATE_n^{cox} &= \mathbb{E}_n^{cox}(Y_1) - \mathbb{E}_n^{cox}(Y_0), \end{aligned}$$

and

$$
\begin{aligned}
\hat{F}^{-1,cox}_{n,Y_0}(0.5) &= \inf\left(y : \hat{F}^{cox,r}_{n,Y_0}(y) \geq 0.5\right), \\
\hat{F}^{-1,cox}_{n,Y_1}(0.5) &= \inf\left(y : \hat{F}^{cox,r}_{n,Y_t}(y) \geq 0.5\right), \\
QTE^{cox}_{n}(0.5) &= \hat{F}^{-1,cox}_{n,Y_1}(0.5) - \hat{F}^{-1,cox}_{n,Y_0}(0.5),
\end{aligned}
$$

where, for $t \in \{0,1\}$, $\hat{F}^{cox,r}_{n,Y_t}(y)$ is the rearrangement of $\hat{F}^{cox}_{n,Y_t}(y)$ if $\hat{F}^{cox}_{n,Y_t}(y)$ is not monotone. A similar approach has been proposed by Chernozhukov et al. (2013) and García-Suaza (2015), in different contexts. It is important to stress that constructing policy evaluation estimators based on the Cox Model as described above depends on functional form restrictions, on top of involving cumbersome calculations. Another important drawback of the "Cox" approach is that the resulting unconditional CDF and quantile functions are not invariant to monotone transformations of the outcome variable. The 2SKM approach, on the other hand, does not suffer from these important limitations.

Finally, "Frandsen" approach relies Frandsen (2015b) proposal. That is, we first estimate the CDF of the potential outcomes Y_0 and Y_1, by

$$
\begin{aligned}
\hat{F}^{\text{Frandsen}}_{n,Y_0}(y) &= \frac{\sum_{i=1}^{n}(1-T_i)\,1\{Q_i \leq y\}\,1\{C_i > y\}}{\sum_{i=1}^{n}(1-T_i)\,1\{C_i > y\}}, \\
\hat{F}^{\text{Frandsen}}_{n,Y_1}(y) &= \frac{\sum_{i=1}^{n}T_i 1\{Q_i \leq y\}\,1\{C_i > y\}}{\sum_{i=1}^{n}T_i 1\{C_i > y\}},
\end{aligned}
$$

and then, with such estimators at hand, we estimate our measures of interest by

$$
\begin{aligned}
\mathbb{E}^{\text{Frandsen}}_{n}(Y_0) &= \int y\hat{F}^{\text{Frandsen}}_{n,Y_0}(dy), \\
\mathbb{E}^{\text{Frandsen}}_{n}(Y_1) &= \int y\hat{F}^{\text{Frandsen}}_{n,Y_1}(dy), \\
ATE^{\text{Frandsen}}_{n} &= \mathbb{E}^{\text{Frandsen}}_{n}(Y_1) - \mathbb{E}^{\text{Frandsen}}_{n}(Y_0),
\end{aligned}
$$

and

$$
\begin{aligned}
\hat{F}^{-1,\text{Frandsen}}_{n,Y_0}(0.5) &= \inf\left(y : \hat{F}^{\text{Frandsen},r}_{n,Y_0}(y) \geq 0.5\right), \\
\hat{F}^{-1,\text{Frandsen}}_{n,Y_1}(0.5) &= \inf\left(y : \hat{F}^{\text{Frandsen},r}_{n,Y_t}(y) \geq 0.5\right), \\
QTE^{\text{Frandsen}}_{n}(0.5) &= \hat{F}^{-1,\text{Frandsen}}_{n,Y_1}(0.5) - \hat{F}^{-1,\text{Frandsen}}_{n,Y_0}(0.5),
\end{aligned}
$$

where, for $t \in \{0,1\}$, $\hat{F}^{\text{Frandsen},r}_{n,Y_t}(y)$ is the rearrangement of $\hat{F}^{\text{Frandsen}}_{n,Y_t}(y)$ if $\hat{F}^{\text{Frandsen}}_{n,Y_t}(y)$ is not monotone. Notice that, in order to compute these estimators, one must always observe the censoring random variable C_i, regardless if observation i is censored or not. Such data requirement puts important restrictions on the applicability of the Frandsen's approach. Furthermore, estimators based on Frandsen (2015b) proposal are not suitable to situations in which covariates plays an important role in the policy evaluation, as in our Designs 3 and 4. The 2SKM approach, on the other hand, does not suffer from these important limitations.

B.2. Further numerical results

In the paper, we compare the 2SKM estimators with those based on the competing methods only in terms of bias. Here, we additionally report finite sample comparisons in terms of root mean square error (RMSE). The numerical results are given in the following table, which follow the same structure as discussed in the paper. All findings are highly consistent with our large-sample theoretical results and the simulation results discussed in the paper.

C. Proofs of the results from Sections 3 and 4

In this section we present the proofs of Proposition 1, Corollary 1 and Theorems 1-3.

Proof of Proposition 1: We first show the identification result for the cumulative hazard. Note that under Assumption 3.1, we have that, for $(y, x, t) \in$

Table 1: Simulated Root Mean Square Error under the unconfoundedness setup

DGP=1

	Not Censored			Censoring=10%					Censoring=30%				
Objects / Estimators	2SKM	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen
$\mathbb{E}(Y_1)$	0.0469	0.0464	0.0456	0.0480	0.1370	0.1725	0.0460	0.0453	0.0668	0.4218	0.5670	0.0558	0.0596
$\mathbb{E}(Y_0)$	0.0451	0.0450	0.0442	0.0463	0.0575	0.0898	0.0455	0.0446	0.0496	0.1235	0.2996	0.0478	0.0480
$F^{-1}_{Y_1}(0.5)$	0.0576	0.0572	0.0559	0.0605	0.1765	0.1444	0.3346	0.0589	0.0678	0.5019	0.5072	0.8648	0.0743
$F^{-1}_{Y_0}(0.5)$	0.0584	0.0585	0.0573	0.0587	0.0542	0.1366	0.0495	0.0566	0.0587	0.0474	0.3942	0.0442	0.0583
ATE	0.0627	0.0624	0.0611	0.0667	0.1116	0.1123	0.0642	0.0631	0.0827	0.3075	0.2789	0.0716	0.0762
$QTE(0.5)$	0.0802	0.0811	0.0795	0.0853	0.1807	0.0854	0.3319	0.0813	0.0894	0.4940	0.1429	0.8440	0.0946

DGP=2

	Not Censored			Censoring=10%					Censoring=30%				
Objects / Estimators	2SKM	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen
$\mathbb{E}(Y_1)$	0.0650	0.0641	0.0625	0.0726	0.1896	0.2666	0.0712	0.0712	0.1138	0.5433	0.8004		0.0926
$\mathbb{E}(Y_0)$	0.0468	0.0462	0.0450	0.0473	0.0573	0.0904	0.0464	0.0454	0.0519	0.1264	0.3035		0.0494
$F^{-1}_{Y_1}(0.5)$	0.0819	0.0805	0.0791	0.0892	0.2338	0.2487	0.4348	0.0900	0.0994	0.5805	0.7911		0.1135
$F^{-1}_{Y_0}(0.5)$	0.0589	0.0565	0.0563	0.0597	0.0558	0.1448	0.0500	0.0572	0.0591	0.0488	0.3959		0.0568
ATE	0.0806	0.0796	0.0774	0.0844	0.1618	0.2000	0.0828	0.0830	0.1241	0.4263	0.5062		0.1055
$QTE(0.5)$	0.1021	0.0989	0.0978	0.1068	0.2323	0.1449	0.4285	0.1071	0.1132	0.5705	0.4088		0.1272

DGP=3

	Not Censored			Censoring=10%					Censoring=30%				
Objects / Estimators	2SKM	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen
$\mathbb{E}(Y_1)$	0.0815	0.0822	0.4848	0.0833	0.2443	0.3747	0.0815	0.4801	0.1300	0.7009	1.1386	0.0979	0.4736
$\mathbb{E}(Y_0)$	0.0555	0.0703	0.2447	0.0621	0.0765	0.1256	0.0708	0.2452	0.0728	0.1838	0.4444	0.0671	0.2462
$F^{-1}_{Y_1}(0.5)$	0.1101	1.0358	0.4900	0.1111	0.2657	0.3856	0.3241	0.4882	0.1264	0.6094	1.1352	0.7762	0.5104
$F^{-1}_{Y_0}(0.5)$	0.0744	0.3995	0.2503	0.0759	0.0708	0.1993	0.4057	0.2495	0.0740	0.0584	0.5595	0.4300	0.2493
ATE	0.0734	0.0756	0.7210	0.0777	0.1953	0.2694	0.0794	0.7154	0.1302	0.5248	0.7028	0.1043	0.7056
$QTE(0.5)$	0.1137	1.4264	0.7262	0.1148	0.2633	0.2195	0.7113	0.7226	0.1280	0.5974	0.5870	0.3783	0.7412

DGP=4

	Not Censored			Censoring=10%					Censoring=30%				
Objects / Estimators	2SKM	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen	2SKM	Ignore	Uncens	Cox	Frandsen
$\mathbb{E}(Y_1)$	0.0942	0.0925	0.4801	0.0948	0.2722	0.4275	0.0931	0.4874	0.1578	0.7768	1.2794	0.1159	0.4703
$\mathbb{E}(Y_0)$	0.0586	0.0718	0.2431	0.0581	0.0740	0.1254	0.0671	0.2457	0.0749	0.1867	0.4489	0.0670	0.2504
$F^{-1}_{Y_1}(0.5)$	0.1181	1.0422	0.4813	0.1202	0.2810	0.4386	0.2977	0.4923	0.1407	0.6350	1.2678	0.8166	0.5062
$F^{-1}_{Y_0}(0.5)$	0.0749	0.3983	0.2500	0.0724	0.0665	0.2035	0.4063	0.2510	0.0747	0.0597	0.5588	0.4279	0.2491
ATE	0.0873	0.0909	0.7139	0.0894	0.2227	0.3204	0.0988	0.7234	0.1609	0.5983	0.8385	0.1264	0.7057
$QTE(0.5)$	0.1245	1.4314	0.7170	0.1273	0.2780	0.2634	0.6799	0.7287	0.1447	0.6232	0.7194	0.4194	0.7349

Note: Simulations based on one thousand Monte Carlo experiments. Sample size equal to 1,000. "2SKM" stands for estimators based on our proposal. "Ignore" stands for estimators based on inverse probability weight (IPW) estimators that ignore the censoring problem. "Uncens" stands for IPW estimators after dropping all censored outcomes. Cox stands for estimators based on the Cox-Proportional hazard model for the treated d and control groups. "Frandsen" stands for the estimators based on Frandsen (2015b).

$(-\infty,\tau_{H_t})\times\mathbb{R}^k\times\{0,1\}$,

$$\begin{aligned}
H_{1t}(y,x) &= \mathbb{P}(Q\le y, X\le x,\delta=1|T=t)\\
&= \mathbb{E}\left(1\{Y\le y\}1\{X\le x\}\mathbb{P}(C\ge Y|X,Y,T)|T=t\right)\\
&= \mathbb{E}\left(1\{Y\le y\}1\{X\le x\}\mathbb{P}(C\ge Y|Y,T)|T=t\right)\\
&= \mathbb{E}\left(1\{Y\le y\}1\{X\le x\}[1-G_t(Y-)]|T=t\right)\\
&= \int_{-\infty}^{y}(1-G_t(\bar{y}-))F_t(d\bar{y},x),
\end{aligned}$$

and

$$\begin{aligned}
1-H_t(y-) &= \mathbb{P}(Q\ge y|T=t)\\
&= \mathbb{P}(C\ge y|T=t)\,\mathbb{P}(Y\ge y|T=t)\\
&= (1-G_t(y-))(1-F_t(y-,\infty)).
\end{aligned}$$

Thus, for $(y,x,t)\in(-\infty,\tau_{H_t})\times\mathbb{R}^k\times\{0,1\}$,

$$\Lambda^{cens}(y,x|T=t)=\int_{-\infty}^{y}\frac{H_{1t}(d\bar{y},x)}{1-H_t(\bar{y}-)}=\int_{-\infty}^{y}\frac{F_t(d\bar{y},x)}{(1-F_t(\bar{y}-,\infty))}=\Lambda(y,x|T=t),$$

as desired.

Next, we show the identification of that $F(y,x,t)$. By the law of total probability and the fact that T is discrete, we have that

$$F(y,x,t)=\sum_{j=0}^{t}\mathbb{P}(T=j)\,\mathbb{P}(Y\le y,X\le x|T=j).$$

Thus, it suffices to show that, for $(y,x,t)\in(-\infty,\tau_{H_t})\times\mathbb{R}^k\times\{0,1\}$,

$$F_t(y,x)=\int_{-\infty}^{y}(1-F_t(\bar{y}-,\infty))\,\Lambda^{cens}(d\bar{y},x|T=t).$$

But we can write $F_t(y,x)$ as

$$\begin{aligned}
F_t(y,x) &= \int_{-\infty}^{y} F_t(dy,x) \\
&= \int_{-\infty}^{y} \left(1-F_t(\bar{y}-,\infty)\right) \frac{F_t(dy,x)}{1-F_t(\bar{y}-,\infty)} \\
&= \int_{-\infty}^{y} \left(1-F_t(\bar{y}-,\infty)\right) \Lambda(d\bar{y},x|T=t).
\end{aligned}$$

Since we already have shown that $\Lambda(y,x|T=t) = \Lambda^{cens}(y,x|T=t)$, this concludes the proof. ■

Proof of Corollary 1: We have shown in Proposition 1 that

$$F(y,x,t) = \sum_{j=0}^{t} \mathbb{P}(T=j) \int_{-\infty}^{y} \left(1-F_j(\bar{y}-,\infty)\right) \Lambda^{cens}(d\bar{y},x|T=j). \tag{C.1}$$

For $t \in \{0,1\}$, one can estimate $\mathbb{P}(T=t)$ by its sample analogue n_t/n, $1-F_t(\bar{y}-,\infty)$ by the time-honored Kaplan and Meier (1958)

$$1-F_{n,t}^{km}(\bar{y}-,\infty) = \prod_{i=1}^{n_t} \left(1-\frac{\delta_{[i:n_t]}}{n_t-i+1}\right)^{1\left\{Q_{i:n_t}\leq y\right\}},$$

and $\Lambda^{cens}(y,x|T=t)$ by (3.2). Given that all these estimators are step functions of the data, we have that, by plugging-in these estimators into (C.1),

$$\begin{aligned}
F_{n,t}^{km}(y,x) &= \sum_{j=0}^{t} \frac{n_j}{n} \int_{-\infty}^{y} \left(1-F_{n,j}^{km}(\bar{y}-,\infty)\right) \Lambda_n^{cens}(d\bar{y},x|T=j) \\
&= \sum_{j=0}^{t} \frac{n_j}{n} \sum_{i=1}^{n_j} \prod_{k=1}^{i-1} \left(1-\frac{\delta_{[k:n_j]}}{n_k-i+1}\right) \frac{\delta_{[i:n_j]} 1\left\{Q_{i:n_j}\leq y\right\} 1\left\{X_{[i:nj]}\leq x\right\}}{n_j-i+1} \\
&= \sum_{j=0}^{t} \sum_{i=1}^{n_j} \left\{\frac{n_j}{n}\frac{\delta_{[i:n_j]}}{n_j-i+1} \prod_{k=1}^{i-1} \left(1-\frac{\delta_{[k:n_j]}}{n_k-i+1}\right)\right\} 1\left\{Q_{i:n_j}\leq y\right\} 1\left\{X_{[i:nj]}\leq x\right\} \\
&= \sum_{j=0}^{t} \sum_{i=1}^{n_t} W_{in_j} 1\left\{Q_{i:n_j}\leq y\right\} 1\left\{X_{[i:nj]}\leq x\right\},
\end{aligned}$$

where

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

which concludes the proof. ■

Proof of Theorem 1: Write

$$\begin{aligned}
S_n^{\varphi}\left(z, \hat{h}_n\right) - S^{\varphi}(z, h_0) &= \int \varphi_{z,h_0}(\bar{y}, \bar{x}, \bar{t}) \left[F_n^{km}(d\bar{y}, d\bar{x}, d\bar{t}) - F(d\bar{y}, d\bar{x}, d\bar{t}) \right] \\
&\quad + \int \left[\varphi_{z,\hat{h}_n}(\bar{y}, \bar{x}, \bar{t}) - \varphi_{z,h_0}(\bar{y}, \bar{x}, \bar{t}) \right] \left(F_n^{km}(d\bar{y}, d\bar{x}, d\bar{t}) - F(d\bar{y}, d\bar{x}, d\bar{t}) \right) \\
&\quad + \int \left[\varphi_{z,\hat{h}_n}(\bar{y}, \bar{x}, \bar{t}) - \varphi_{z,h_0}(\bar{y}, \bar{x}, \bar{t}) \right] F(d\bar{y}, d\bar{x}, d\bar{t}) \\
&= A_{1n}\left(z, \hat{h}_n\right) + A_{2n}(z, h_0) + A_{3n}(z, h_0)
\end{aligned}$$

From Stute (1993), we have that under Assumptions 3.1 and 4.3, $\int \varphi_{z,h_0}(\bar{y}, \bar{x}, \bar{t}) F_n^{km}(\bar{y}, \bar{x}, \bar{t})$ is Glivenko-Cantelli. Thus,

$$\sup_{z \in \mathcal{W}} |A_{1_n}(z, h_0)| = o_{\mathbb{P}}(1). \tag{C.2}$$

Next, by Assumptions 4.1, 4.2 and 4.3, we have

$$\sup_{z \in \mathcal{W}} \left| A_{2_n}\left(z, \hat{h}_n\right) \right| = o_{\mathbb{P}}(1), \tag{C.3}$$

$$\sup_{z \in \mathcal{W}} |A_{3n}(z, h_0)| = \sup_{z \in \mathcal{W}} \left| S^{\varphi}\left(z, \hat{h}_n\right) - S^{\varphi}(z, h_0) \right| = o_{\mathbb{P}}(1). \tag{C.4}$$

Thus, by triangle inequality and (C.2)-(C.4), the proof is completed. ■

Proof of Theorem 2: First, we establish the linear representation of $\sqrt{n}\left(S_n^{\varphi}(z, h_n) - S^{\varphi}(z, h_0)\right)$. For this end, write

$$
\begin{aligned}
\sqrt{n}\left(S_n^{\varphi}\left(z,\hat{h}_n\right)-S^{\varphi}\left(z,h_0\right)\right) &= \sqrt{n}\int\varphi_{z,h_0}\left(\bar{y},\bar{x},\bar{t}\right)\left[F_n^{km}\left(d\bar{y},d\bar{x},d\bar{t}\right)-F\left(d\bar{y},d\bar{x},d\bar{t}\right)\right] \\
&\quad+\sqrt{n}\int\left[\varphi_{z,\hat{h}_n}\left(\bar{y},\bar{x},\bar{t}\right)-\varphi_{z,h_0}\left(\bar{y},\bar{x},\bar{t}\right)\right]\left(F_n^{km}\left(d\bar{y},d\bar{x},d\bar{t}\right)-F\left(d\bar{y},d\bar{x},\right.\right. \\
&\quad+\sqrt{n}\int\left[\varphi_{z,\hat{h}_n}\left(\bar{y},\bar{x},\bar{t}\right)-\varphi_{z,h_0}\left(\bar{y},\bar{x},\bar{t}\right)\right]F\left(d\bar{y},d\bar{x},d\bar{t}\right) \\
&= \mathbb{G}_n^{km}\left(\varphi_{z,h_0}\right)+\mathbb{G}_n^{km}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right)+\sqrt{n}\mathbb{P}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right).
\end{aligned}
$$

From Stute (1996), we have that, under Assumptions 3.1, 4.6, 4.8 and 4.9, and Conditions 1 or 2, uniformly in $z\in\mathcal{W}$

$$
\mathbb{G}_n^{km}\left(\varphi_{z,h_0}\right)-n^{-1/2}\sum_{i=1}^{n}\eta_i^{\varphi}\left(z,h_0\right)=o_{\mathbb{P}}\left(1\right), \tag{C.5}
$$

where $\eta_i^{\varphi}\left(z,h_0\right)$ is defined as in (4.1). From van der Vaart and Wellner (2007), we have that under Assumptions 4.4, 4.5, and 4.6,

$$
\mathbb{G}_n^{km}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right)=o_{\mathbb{P}}\left(1\right). \tag{C.6}
$$

Finally note that we can rewrite $\sqrt{n}\mathbb{P}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right)$ as

$$
\begin{aligned}
\sqrt{n}\mathbb{P}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right) = \sqrt{n}\left(S^{\varphi}\left(z,\hat{h}_n\right)-S^{\varphi}\left(z,h_0\right)-\Gamma^{\varphi}\left(z,h_0\right)[\hat{h}_n-h_0]\right) \\
+\sqrt{n}\Gamma^{\varphi}\left(z,h_0\right)[\hat{h}_n-h_0].
\end{aligned}
$$

Thus, from Assumptions 4.4, 4.5 and 4.7, uniformly in $z\in\mathcal{W}$,

$$
\sqrt{n}\mathbb{P}\left(\varphi_{z,\hat{h}_n}-\varphi_{z,h_0}\right)-n^{-1/2}\sum_{i=1}^{n}\kappa_i^{\varphi}\left(z,h_0\right)=o_{\mathbb{P}}\left(1\right). \tag{C.7}
$$

Then, putting (C.5)-(C.7) together, we have that

$$\begin{aligned}\sqrt{n}\left(S_n^{\varphi}\left(z,\hat{h}_n\right)-S^{\varphi}\left(z,h_0\right)\right) &= n^{-1/2}\sum_{i=1}^{n}\left[\eta_i^{\varphi}\left(z,h_0\right)+\kappa_i^{\varphi}\left(z,h_0\right)\right]+R_{1n},\\ &= n^{-1/2}\sum_{i=1}^{n}\Psi_i^{\varphi}\left(z\right)+R_{1n}\end{aligned}$$

where

$$\sup_{z\in\mathcal{W}}\left|R_{1n}\right|=o_{\mathbb{P}}\left(1\right)\ \text{ and }\ \Psi_i^{\varphi}\left(z\right)\equiv\eta_i^{\varphi}\left(z,h_0\right)+\kappa_i^{\varphi}\left(z,h_0\right).$$

Next, we establish the weak convergence of $\sqrt{n}\left(S_n^{\varphi}\left(z,\hat{h}_n\right)-S^{\varphi}\left(z,h_0\right)\right)$. From its linear representation, it is enough to prove that weak convergence of the dominant term $\Psi^{\varphi}\left(z\right)$. This will be done by showing that the class of functions $\left\{\Psi^{\varphi}\left(z\right)\right\}$ is Donsker. But from Theorem 2.10.6 in van der Vaart and Wellner (1996) it suffices to show that the classes of functions $\left\{\varphi_{z,h_0}\right\}$, $\left\{\gamma_{0t}\right\}$, $\left\{\gamma_{j,t,z,h_0}^{\varphi}\right\}$ $(j=\{1,2\}$ and $t\in\{0,1\})$, and $\left\{\kappa^{\varphi}\left(z,h_0\right)\right\}$ are Donsker.

First, $\left\{\varphi_{z,h_0}\right\}$ and $\left\{\kappa^{\varphi}\left(z,h_0\right)\right\}$are Donsker by Assumptions 4.6 and 4.7, respectively. For $t\in\{0,1\}$, the function γ_{0t} does not depend on z nor on h, and so it is clearly Donsker. Next, consider $\gamma_{1,t,z,h_0}^{\varphi}$. By Theorem 2.5.6 in van der Vaart and Wellner (1996) it suffices to show that

$$\int_0^{\infty}\sqrt{\ln N_{[]}\left(\varepsilon,\left\{\gamma_{1,0,z,h_0}^{\varphi}\right\},L_2\left(H\right)\right)}d\varepsilon<\infty, \tag{C.8}$$

where $N_{[\,]}$ is the bracketing number, H is the probability measure corresponding to the joint distribution of (Q,δ,X,T), and $L_2\left(H\right)$ is the L_2-norm. Fix $\varepsilon>0$. For $t\in\{0,1\}$, write

$$\begin{aligned}\gamma_{1,t,z,h_0}^{\varphi}\left(Q\right)=&\frac{1}{1-H_t\left(Q\right)}\int 1\left\{Q<\bar{w}\right\}\varphi_{z,h}^{+}\left(\bar{w},\bar{x},t\right)\gamma_{0t}\left(\bar{w}\right)H_t\left(d\bar{w},d\bar{x}\right)\\ &+\frac{1}{1-H_t\left(Q\right)}\int 1\left\{Q<\bar{w}\right\}\varphi_{z,h}^{-}\left(\bar{w},\bar{x},t\right)\gamma_{0t}\left(\bar{w}\right)H_t\left(d\bar{w},d\bar{x}\right),\end{aligned} \tag{C.9}$$

where

$$\begin{aligned}\varphi^{+}_{z,h}(Q,X,t) &= \varphi_{z,h}(Q,X,t)\,1\left\{\varphi_{z,h}(Q,X,t)\geq 0\right\},\\ \varphi^{-}_{z,h}(Q,X,t) &= \varphi_{z,h}(Q,X,t)\,1\left\{\varphi_{z,h}(Q,X,t)<0\right\}.\end{aligned}$$

From (C.9), we have that for $t\in\{0,1\}$, $\gamma^{\varphi}_{1,t,z,h_0}(Q)$ is the sum of two monotone uniformly bounded class of functions, and hence, by Theorem 2.7.5 in van der Vaart and Wellner (1996), the integral in (C.8) is finite. Following the same steps, we have that $\left\{\gamma^{\varphi}_{2,t,z,h_0}\right\}$ is also Donsker, concluding our proof. ■

Proof of Corollary 2: it follows from Theorem 2 and the functional delta method, cf. Theorem 3.9.4 in van der Vaart and Wellner (1996).■

Proof of Theorem 3: In order to prove the bootstrap validity, we will use similar steps as in Chen et al. (2003). Denote $v^{*}_{n}(z,h)=\sqrt{n}\left(S^{*,\varphi}_{n}(z,h)-S^{\varphi}_{n}(z,h)\right)$ and $v_n(z,h)=\sqrt{n}\left(S^{\varphi}_{n}(z,h)-S^{\varphi}(z,h)\right)$. By the triangle inequality, and Assumption 4.6,

$$\text{(C.10)}\qquad \sup_{(z,h),(z,h')\in\mathcal{W}\times H_\delta}\left|v^{*}_{n}\left(z,h'\right)-v^{*}_{n}(z,h)\right| = o_{\mathbb{P}^{*}}(1),$$

$$\text{(C.11)}\qquad \sup_{(z,h),(z,h')\in\mathcal{W}\times H_\delta}\left|v_{n}\left(z,h'\right)-v_{n}(z,h)\right| = o_{\mathbb{P}^{*}.}(1).$$

Next, by adding and subtracting terms, we have that, uniformly in $z\in\mathcal{W}$,

$$\begin{aligned} &\sqrt{n}\left[S^{*,\varphi}_{n}\left(z,\hat{h}^{*}_{n}\right)-S^{\varphi}_{n}\left(z,\hat{h}_{n}\right)\right]\\ =\;&\sqrt{n}\left(\left[S^{*,\varphi}_{n}(z,h_0)-S^{\varphi}_{n}(z,h_0)\right]+\Gamma^{\varphi}\left(z,\hat{h}_{n}\right)\left[\hat{h}^{*}_{n}-\hat{h}_{n}\right]\right)\\ &+\sqrt{n}\left(\left[S^{*,\varphi}_{n}\left(z,\hat{h}_{n}\right)-S^{\varphi}_{n}\left(z,\hat{h}_{n}\right)\right]-\left[S^{*,\varphi}_{n}(z,h_0)-S^{\varphi}_{n}(z,h_0)\right]\right)\\ &+\sqrt{n}\left(\left[S^{*,\varphi}_{n}\left(z,\hat{h}^{*}_{n}\right)-S^{\varphi}_{n}\left(z,\hat{h}^{*}_{n}\right)\right]-\left[S^{*,\varphi}_{n}\left(z,\hat{h}_{n}\right)-S^{\varphi}_{n}\left(z,\hat{h}_{n}\right)\right]\right)\\ &+\sqrt{n}\left(\left[S^{\varphi}_{n}\left(z,\hat{h}^{*}_{n}\right)-S^{\varphi}_{n}\left(z,\hat{h}_{n}\right)\right]-\left[S^{\varphi}\left(z,\hat{h}^{*}_{n}\right)-S^{\varphi}\left(z,\hat{h}_{n}\right)\right]\right)\\ &+\sqrt{n}\left(\left[S^{\varphi}\left(z,\hat{h}^{*}_{n}\right)-S^{\varphi}\left(z,\hat{h}_{n}\right)\right]-\Gamma^{\varphi}\left(z,\hat{h}_{n}\right)\left[\hat{h}^{*}_{n}-\hat{h}_{n}\right]\right)\\ \text{(C.12)}\quad =\;&\sqrt{n}\left(S^{*,\varphi}_{n}(z,h_0)-S^{\varphi}_{n}(z,h_0)\right)+\sqrt{n}\left(\Gamma^{\varphi}\left(z,\hat{h}_{n}\right)\left[\hat{h}^{*}_{n}-h_{n}\right]\right)+o_{\mathbb{P}^{*}}(1)\end{aligned}$$

where the second and third term are $o_{\mathbb{P}^{*}}(1)$ from the stochastic equicontinuity property

(C.10), the fourth term is $o_{\mathbb{P}^*}(1)$ from (C.11), and the fifth term from Assumption 4.5 and 4.10.

From Assumptions 4.7 and 4.11, the Donsker results derived in Theorem 2, and Theorem 3.6.1 in van der Vaart and Wellner (1996), it follows that,

$$\sqrt{n}\left(S_n^{*,\varphi}(z,h_0)-S_n^{\varphi}(z,h_0)\right)+\sqrt{n}\left(\Gamma^{\varphi}\left(z,\hat{h}_n\right)[\hat{h}_n^*-h_n]\right)+o_{\mathbb{P}^*}(1)\underset{*}{\Rightarrow}\mathbb{G}. \tag{C.13}$$

Thus, combining (C.12) with (C.13) concludes our proof. ■

Proof of Corollary 3: it follows from Theorem 3 and the functional delta method for the bootstrap, cf. Theorem 3.9.11 in van der Vaart and Wellner (1996).■

D. Proofs of the Propositions in Section 5

D.1. Proof of Proposition 2

In this section we present the proof of Proposition 2. Denote the true propensity score by $p_0(X)$. Next, we verify the high level conditions of Theorems 2 and 3 using different estimators $p_0(X)$. Note that the conditions for applying Theorem 2 are stronger than those for of Theorems 1, so uniform consistency will follow from the conditions of Theorems 2.

We first establish the weak convergence for average, distributional, and quantile treatment effects estimators. To this end, we will verify the conditions for applying Theorem 2 for the functions

$$\varphi_p^{ate,1}=\frac{TQ}{p(X)},\qquad \varphi_p^{ate,0}=\frac{(1-T)Q}{1-p(X)},$$
$$\varphi_{y,p}^{dte,1}=\frac{T1\{Y\leq y\}}{p(X)},\qquad \varphi_{y,p}^{dte,0}=\frac{(1-T)1\{Y\leq y\}}{1-p(X)}.$$

Here, we allow one to use parametric or nonparametric estimators for p_0, as long as Assumptions A.1-A.3 are satisfied.

For all functions (and all three propensity score estimators), Assumption 4.4 follow from Lemmas 1 and 2; Assumption 4.5 follows from Lemmas 3 and 4; Assumption 4.6 follows from Lemmas 5 and 6; Assumption 4.7 follows from Lemmas 7-10; and Assumptions 4.8 and 4.9 are satisfied by the regularity conditions imposed in the proposition.

First, let's focus on the average treatment effect estimator ATE_n^{km}. Following the notation in Theorem 2, define

$$\begin{aligned}\alpha_n^{ate} &= \left(\hat{S}_n^{\varphi^{ate,0}}(\hat{p}_n), \hat{S}_n^{\varphi^{ate,1}}(\hat{p}_n)\right)', \\ \alpha^{ate} &= \left(S^{\varphi^{ate,0}}(p_0), S^{\varphi^{ate,1}}(p_0)\right)'.\end{aligned}$$

Then, by Theorem 2 and Example 19.18 in van der Vaart (1998), we have

$$\sqrt{n}\left(\boldsymbol{\alpha}_n^{ate} - \boldsymbol{\alpha}^{ate}\right) \xrightarrow{d} N\left(0, \Omega^{ate}\right) \tag{D.1}$$

where $\Omega^{ate} = E\left[\Psi_{ate}\Psi_{ate}'\right]$, $\Psi_{ate} = (\psi_0, \psi_1)'$ with

$$\psi_t = \eta_t^{ate}(p_0) - \mathbb{E}(Y_t) + \kappa_t^{ate}(p_0),$$

$t \in \{0,1\}$, and $\eta_0^{ate}(p_0)$ and $\eta_1^{ate}(p_0)$ are defined as in (4.1) with $\varphi_{p_0}^{ate,0}(Q,X,T)$ and $\varphi_{p_0}^{ate,1}(Q,X,T)$, respectively, and $\kappa_t^{ate}(p_0)$, $t \in \{0,1\}$, are defined as in Lemma 7 if one adopts a parametric approach, or in Lemma 9 if one uses nonparametric methods to estimate p_0. Thus, the asymptotic normality of the average treatment effect estimator ATE_n^{km} follows an standard application of the Delta method to (D.1).

Now, we move to the distributional treatment effect estimator $DTE_n^{km}(\cdot)$. Define

$$\begin{aligned}\alpha_n^{dte} &= \left(\hat{S}_n^{\varphi^{dte,0}}(y_0, \hat{p}_n), \hat{S}_n^{\varphi^{dte,1}}(y_1, \hat{p}_n)\right)', \\ \alpha^{dte} &= \left(S^{\varphi^{dte,0}}(y_0, p_0), S^{\varphi^{dte,1}}(y_1, p_0)\right)',\end{aligned}$$

and $\nu = (y_0, y_1)' \in Y_H \times Y_H$, where $Y_H \equiv Y \cap (-\infty, \tau_H)$. Then, by Theorem 2 and Example 19.18 in van der Vaart (1998), we have

$$\sqrt{n}\left(\alpha_n^{dte}(\cdot) - \alpha^{dte}(\cdot)\right) \Rightarrow \mathbb{G}^{dte}(\cdot) \text{ in } l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H), \tag{D.2}$$

where $G^{dte}(v) = \left(\mathbb{G}_0^{dte}(y_0), \mathbb{G}_1^{dte}(y_1)\right)'$ is a two-dimensional mean zero Gaussian process with covariance function $\Omega^{dte}(y_0, y_1) = E\left[\Psi_{dte}(\nu_0)\Psi_{dte}(\nu_1)'\right]$, $\Psi_{dte}(\nu) = (\psi_0(y_0), \psi_1(y_1))'$ with, for $t \in \{0,1\}$

$$\psi_t(y) = \eta_t^{dte}(y, p_0) - F_{Y_t}(y) + \kappa_t^{dte}(y, p_0),$$

$t \in \{0,1\}$, and $\eta_0^{dte}(y,p_0)$ and $\eta_1^{dte}(y,p_0)$ are defined as in (4.1) with $\varphi_{y,p_0}^{dte,0}(Q,X,T)$ and $\varphi_{y,p_0}^{dte,1}(Q,X,T)$, respectively, and $\kappa_t^{dte}(y,p_0)$, $t \in \{0,1\}$, are defined as in Lemma 8 if one adopts a parametric approach, or in Lemma 10 if one uses nonparametric methods to estimate p_0. Thus, the weak convergence of the Distributional treatment effect estimator $DTE_n^{km}(\cdot)$ follows from an standard application of the functional Delta method to (D.2).

Next, we move to the quantile treatment effect estimator $QTE_n^{km}(\cdot)$. Define

$$\begin{aligned}\alpha_n^{qte} &= \left(\hat{F}_{n,Y_0}^{km,-1}, \hat{F}_{n,Y_1}^{km,-1}\right)', \\ \alpha^{qte} &= \left(F_{Y_0}^{-1}, F_{Y_1}^{-1}\right)',\end{aligned}$$

and $\omega = (\tau_0, \tau_1)' \in (0,\bar{\tau}) \times (0,\bar{\tau})$, where

$$\begin{aligned}\hat{F}_{n,Y_0}^{km,-1}(\tau) &= \inf\left\{y : \hat{F}_{n,Y_0}^{km,r}(y) \geq \tau\right\}, \\ \hat{F}_{n,Y_1}^{km,-1}(\tau) &= \inf\left\{y : \hat{F}_{n,Y_1}^{km,r}(y) \geq \tau\right\},\end{aligned}$$

and, for $t \in \{0,1\}$, $\hat{F}_{n,Y_0}^{km,r}(y)$ is the rearrangement of $\hat{F}_{n,Y_t}^{km}(y)$ as proposed by Chernozhukov et al. (2010), with $\hat{F}_{n,Y_t}^{km,r}(y) = 1$ if $\hat{F}_{n,Y_t}^{km}(y) > 1$, and $\hat{F}_{n,Y_t}^{km,r}(y) = 0$ if $\hat{F}_{n,Y_t}^{km}(y) < 0$. This transformation guarantees that $\hat{F}_{n,Y_t}^{km,r}(y)$ is a proper CDF. Given that we assume that F_{Y_t} is continuously differentiable with strictly positive derivative everywhere, we have, from Chernozhukov et al. (2010), that, for $t \in \{0,1\}$,

$$\sqrt{n}\left(\hat{F}_{n,Y_t}^{km,r} - F_{Y_t}\right)(\cdot) = \sqrt{n}\left(\hat{F}_{n,Y_t}^{km} - F_{Y_t}\right) + o_{\mathbb{P}}(1), \tag{D.3}$$

implying that the rearranged 2SKM estimator is first order equivalent to the original 2SKM estimator, and thus inherits the limit distribution. Then, from the fact that $\hat{F}_{n,Y_t}^{km,r}$ is a proper distribution function, and that the quantile mapping is Hadamard differentiable, cf. Lemma 3.9.23 in van der Vaart and Wellner (1996), we can combine (D.2) with (D.3) and the functional delta method, cf. Theorem 3.9.4 in van der Vaart and Wellner (1996), and get that

$$\sqrt{n}\left(\alpha_n^{qte}(\cdot) - \alpha^{qte}(\cdot)\right) \Rightarrow \mathcal{Q}(\cdot) \;\; in \; l^{\infty}((0,\bar{\tau})) \times l^{\infty}((0,\bar{\tau})) \tag{D.4}$$

where $Q(\omega) = (\mathcal{Q}_0(\tau_0), \mathcal{Q}_1(\tau_1))$ is a two-dimensional mean zero Gaussian process such that

$$\mathcal{Q}_t(\tau_t) \equiv -\frac{\mathbb{G}_t^{dte}\left(F_{Y_t}^{-1}(\tau_t)\right)}{f\left(F_{Y_t}^{-1}(\tau_t)\right)},\ t \in \{0,1\}.$$

The weak convergence of the quantile treatment effects estimator $QTE_n^{km}(\cdot)$ in follows from an standard application of the functional delta method to (D.4).

From (D.1)-(D.4) and the functional delta method, we have that all our treatment effect estimators converge weakly to a tight mean zero Gaussian process.

To conclude the proof, we need to show that the validity of the bootstrap. This follows from verifying conditions 4.10 and 4.11 for our functionals, and then applying the functional delta method for the bootstrap, cf. Theorem 3.9.11 in van der Vaart and Wellner (1996). For the parametric propensity score case, condition 4.10 follows from arguments similar to those of Lemma 1, Assumption A.1 and Theorem 3.6.1 in van der Vaart and Wellner (1996), whereas for the nonparametric case, condition 4.10 follows from arguments similar to those of Lemma 2, and the fact we are "undersmoothing", see the discussion in Chapter 4.4.2 of Hall (1992). Condition 4.11 follows from Lemmas 7-10 and similar arguments to those presented in Lemmas 3 and 4. ■

D.2. *Proof of Proposition 3*

In this section we present the proof of Proposition 3. Such proof consist of verifying the high level conditions of Theorems 1-3. Denote the true $\mathbb{P}(Z=1|X)$ by $e_0(X)$. Similar to Proposition 2, we establish the weak convergence for local average, distributional, and quantile treatment effects estimators by verifying the conditions for applying Theorem 2 for the functions

$$\varphi_e^{late,11} = \frac{TZQ}{e(X)}, \qquad \varphi_e^{late,10} = \frac{T(1-Z)Q}{1-e(X)},$$
$$\varphi_e^{late,01} = \frac{(1-T)ZQ}{e(X)}, \qquad \varphi_e^{late,00} = \frac{(1-T)(1-Z)Q}{1-e(X)},$$
$$\varphi_{y,e}^{ldte,11} = \frac{TZ1\{Q \le y\}}{e(X)}, \qquad \varphi_{y,e}^{ldte,10} = \frac{T(1-Z)1\{Q \le y\}}{1-e(X)},$$
$$\varphi_{y,e}^{ldte,01} = \frac{(1-T)Z1\{Q \le y\}}{e(X)}, \qquad \varphi_{y,e}^{ldte,00} = \frac{(1-T)(1-Z)1\{Q \le y\}}{1-e(X)},$$

where we allow one to use parametric or nonparametric estimators for e_0, as long as

Assumptions A.1-A.3 are satisfied.

Following the notation in Theorem 2, denote

$$\begin{aligned}
\alpha_n^{late} &= \left(\hat{S}_n^{\varphi^{late,00}}(\hat{e}_n), \hat{S}_n^{\varphi^{late,01}}(\hat{e}_n), \hat{S}_n^{\varphi^{late,10}}(\hat{e}_n), \hat{S}_n^{\varphi^{late,11}}(\hat{e}_n)\right)', \\
\alpha^{late} &= \left(S^{\varphi^{late,00}}(e_0), S^{\varphi^{late,01}}(e_0), S^{\varphi^{late,10}}(e_0), S^{\varphi^{late,11}}(e_0)\right)', \\
\alpha_n^{ldte} &= \left(\hat{S}_n^{\varphi^{ldte,00}}(y,\hat{e}_n), \hat{S}_n^{\varphi^{ldte,01}}(y,\hat{e}_n), \hat{S}_n^{\varphi^{ldte,10}}(y,\hat{e}_n), \hat{S}_n^{\varphi^{ldte,11}}(y,\hat{e}_n)\right)', \\
\alpha^{ldte} &= \left(S^{\varphi^{ldte,00}}(y,e_0), S^{\varphi^{ldte,01}}(y,e_0), S^{\varphi^{ldte,10}}(y,e_0), S^{\varphi^{ldte,11}}(y,e_0)\right)'.
\end{aligned}$$

Thus, from arguments similar to those in Proposition 2, we have that

$$\sqrt{n}\left(\boldsymbol{\alpha}_n^{late} - \boldsymbol{\alpha}^{late}\right) \xrightarrow{d} N\left(0, \Omega^{late}\right) \tag{D.5}$$

where $\Omega^{late} = \mathbb{E}\left[\Psi_{late}\Psi_{late}'\right]$, $\Psi_{ate} = \left(\psi_{00}, \psi_{01}, \psi_{10,}\psi_{11}\right)'$ with

$$\psi_{tz} = \eta_{tz}^{late}(e_0) - S^{\varphi^{late,td}}(e_0) + \kappa_{tz}^{late}(e_0),$$

$t, z \in \{0,1\}$, and the other terms are defined analogously to those in Proposition 2. The weak convergence of the $LATE_n^{km}$ follows from the fact that, for $t \in \{0,1\}$,

$$\hat{\kappa}_{t,n}(\hat{e}_n) = \kappa_t(e_0) + o_{\mathbb{P}}(1), \tag{D.6}$$

and the delta method applied to (D.5).

Analogously, we have

$$\sqrt{n}\left(\alpha_n^{ldte}(\cdot) - \alpha^{ldte}(\cdot)\right) \Rightarrow \mathbb{G}^{ldte}(\cdot) \text{ in } l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H), \tag{D.7}$$

where $\mathbb{G}^{ldte}(\cdot) = \left(\mathbb{G}_{00}^{ldte}(y_{00}), \mathbb{G}_{01}^{ldte}(y_{01}), \mathbb{G}_{10}^{ldte}(y_{10}), \mathbb{G}_{11}^{ldte}(y_{11})\right)'$ is a four-dimensional mean zero Gaussian process with covariance function $\Omega^{ldte}(y_{00}, y_{01}, y_{10,}, y_{11}) = \mathbb{E}\left[\Psi_{ldte}(\nu_0)\Psi_{ldte}(\nu_1)'\right]$,

$$\Psi_{ldte}(\nu) = \left(\psi_{00}(y_{00}), \psi_{01}(y_{01}), \psi_{10}(y_{10}), \psi_{11}(y_{11})\right)'$$

with

$$\psi_{tz}(y) = \eta_{tz}^{ldte}(y,e_0) - S^{\varphi^{late,td}}(y,e_0) + \kappa_{td}^{ldte}(y,e_0),$$

$t,d=0,1$, and the other terms are defined analogously to those in Proposition 2. The weak convergence of the $LDTE_n^{km}(\cdot)$ follows from (D.6) and the functional delta method.

As in Proposition 2, the weak convergence of the $LQTE_n^{km}(\cdot)$ follows from (D.6), (D.7), the rearrangement steps based on Chernozhukov et al. (2010), and the functional delta method. The validity of the bootstrap follows from same arguments as in Proposition 2 and is therefore omitted.

D.3. *Proof of Proposition 4*

In this section we present the proof of Proposition 4. Because $n_{gj}/n \to \alpha_{gj}$, with $0<\alpha_{gj}<1$, any term that is $O_{\mathbb{P}}\left(n_{gj}^{-\omega}\right)$ is also $O_{\mathbb{P}}\left(n^{-\omega}\right)$; similarly term that are $o_{\mathbb{P}}\left(n_{gj}^{-\omega}\right)$ are also $o_{\mathbb{P}}\left(n^{-\omega}\right)$.

The proof of Proposition 4 follows a different path than previous proofs: instead of directly verifying the Assumptions of Theorems 2, we show that the changes-in-changes treatment effects parameters can be written as Hadamard differentiable functions of $\left(\hat{F}^{km}_{n,Y_{00}}\left(y_{00}\right),\hat{F}^{km}_{n,Y_{01}}\left(y_{01}\right),\hat{F}^{km}_{n,Y_{10}}\left(y_{10}\right),\hat{F}^{km}_{n,Y_{11}}\left(y_{11}\right)\right)$. Then, the the weak convergence and the bootstrap validity follows from the functional delta method.

To this end, denote

$$
\begin{aligned}
\alpha_n^{cic}\left(y\right) &= \left(\hat{F}^{km}_{n,Y_{00}}\left(y_{00}\right),\hat{F}^{km}_{n,Y_{01}}\left(y_{01}\right),\hat{F}^{km}_{n,Y_{10}}\left(y_{10}\right),\hat{F}^{km}_{n,Y_{11}}\left(y_{11}\right)\right)',\\
\alpha^{cic}\left(y\right) &= \left(F_{Y_{00}}\left(y_{00}\right),F_{Y_{01}}\left(y_{01}\right),F_{Y_{10}}\left(y_{10}\right),F_{Y_{11}}\left(y_{11}\right)\right)'
\end{aligned}
$$

and $y=\left(y_{00},y_{01},y_{10},y_{11}\right)'\in\mathcal{W}_H\equiv\mathcal{Y}_H\times\mathcal{Y}_H\times\mathcal{Y}_H\times\mathcal{Y}_H$, where $\mathcal{Y}_H\equiv\mathcal{Y}\cap(-\infty,\tau_H^{cic}]$. Under Condition 1 and Assumption 4.9[12], we have that, from Stute (1995),

$$
\alpha_n^{cic}\left(y\right)-\alpha^{cic}\left(y\right)=\frac{1}{n}\sum_{i=1}^{n}\Psi_i^{cic}\left(y\right)+o_{\mathbb{P}}\left(n^{-1/2}\right) \tag{D.8}
$$

where $\Psi_i^{cic}\left(y\right)=\left(\eta_i^{cic}\left(y\right)'-\alpha^{cic}\left(y\right)\right)$,

$$
\eta_i^{cic}\left(y\right)=\left(\alpha_{00}^{-1}\eta^{cic}_{00,\ i}\left(y_{00}\right),\alpha_{01}^{-1}\eta^{cic}_{01,\ i}\left(y_{01}\right),\alpha_{10}^{-1}\eta^{cic}_{10,\ i}\left(y_{10}\right),\alpha_{11}^{-1}\eta^{cic}_{11,\ i}\left(y_{11}\right)\right)'
$$

12 If Condition 1 does not hold, we need to set $\mathcal{Y}_H\equiv\mathcal{Y}\cap(-\infty,\tau_H)$. If this is the case, we can drop Assumption 4.9.

and, for $g, j = 0, 1$,

$$\eta^{cic}_{gj,\ i}(w) = 1\{G_i = g, I_i = j\}[1\{Q_i \le w\}\gamma_{0,gj}(Q_i)\delta + \gamma_{1,gj,w}(Q_i)(1-\delta_i) - \gamma_{2,gj,w}(Q_i)],$$

with

$$\begin{aligned}
\gamma_{0,gj}(y) &= \exp\left\{\int_0^{y-} \frac{H_{0,gj}(d\bar{w})}{1 - H_{gj}(\bar{w})}\right\},\\
\gamma_{1,gj,w}(y) &= \frac{1}{1 - H_{gj}(y)}\int 1\{y < \bar{w}\}1\{\bar{w} \le w\}\gamma_{0,gj}(\bar{w})H_{gj}(d\bar{w}, d\bar{x}),\\
\gamma_{2,gj,w}(y) &= \int\int \frac{1\{\bar{v} < y, \bar{v} < \bar{w}\}1\{\bar{w} \le w\}}{[1 - H_{gj}(\bar{v})]^2}\gamma_{0,gj}(\bar{w})H_{0,gj}(d\bar{v})H_{1,gj}(d\bar{w}, d\bar{x}),
\end{aligned}$$

and

$$\begin{aligned}
H_{gt}(w) &= \mathbb{P}(Q \le w | G = g, I = j),\\
H_{1,gt}(w, x) &= \mathbb{P}(Q \le w, X \le x, \delta = 1 | G = g, I = j),\\
H_{0,gj}(w) &= \mathbb{P}(Q \le w, \delta = 0 | G = g, I = j).
\end{aligned}$$

From the asymptotic linear representation in (D.8), and the fact that the class of function $\{1\{Y \le w\} : w \in \mathbb{R}\}$ is Donsker, we have that

$$\text{(D.9)}\quad \sqrt{n}\left(\alpha^{cic}_n(\cdot) - \alpha^{cic}(\cdot)\right) \Rightarrow \mathbb{G}^{cic}(\cdot) \text{ in } l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H) \times l^\infty(\mathcal{Y}_H),$$

where $G^{cic}(\cdot) = \left(\mathbb{G}^{cic}_{00}(y_{00}), \mathbb{G}^{cic}_{01}(y_{01}), \mathbb{G}^{cic}_{10}(y_{10}), \mathbb{G}^{cic}_{11}(y_{11})\right)'$ is a four-dimensional mean zero Gaussian process with covariance function $\Omega^{cic}(y_{00}, y_{01}, y_{10,}, y_{11}) = E\left[\Psi^{cic}(y_0)\Psi_{cic}(y_1)'\right]$.

Next, by the functional delta method and the Hadamard differentiability of the probability-probability transformation $F^{-1}_{Y_{00}}(F_{Y_{01}}(\cdot))$, see Lemma A.1 in Callaway et al. (2015), we have that,

$$\sqrt{n}\left(\hat{F}^{km,-1}_{n,Y_{00}}\left(\hat{F}^{km}_{nY_{01}}(\cdot)\right) - F^{-1}_{Y_{00}}(F_{Y_{01}}(\cdot))\right) \Rightarrow \check{\mathbb{G}}_{00}(\cdot) \equiv \frac{\mathbb{G}^{cic}_{01}(y) - \mathbb{G}^{cic}_{00}\left(F^{-1}_{Y_{00}}(F_{Y_{01}}(\cdot))\right)}{f_{00}\left(F^{-1}_{Y_{00}}(F_{Y_{01}}(\cdot))\right)} \text{ in } l^\infty(\mathcal{Y}_H).$$

By the functional delta method and the Hadamard differentiability of the quantile-

quantile transformation $F_{Y_{10}}\left(F_{Y_{00}}^{-1}(\cdot)\right)$, see Problem 3.9.3 in van der Vaart and Wellner (1996), we obtain

(D.10)
$$\sqrt{n}\left(\hat{F}_{n,Y_{10}}^{km}\left(\hat{F}_{n,Y_{00}}^{km,-1}\left(\hat{F}_{nY_{01}}^{km}(\cdot)\right)\right)-F_{10}\left(F_{Y_{00}}^{-1}\left(F_{Y_{01}}(\cdot)\right)\right)\right)\Rightarrow\tilde{\mathbb{G}}_{10}(\cdot)\ \text{ in } l^{\infty}\left(\mathcal{Y}_H\right),$$

where
$$\tilde{\mathbb{G}}_{10}(\cdot)\equiv\mathbb{G}_{10}^{cic}\left(F_{Y_{00}}^{-1}\left(F_{Y_{01}}(\cdot)\right)\right)-\frac{f_{10}\left(F_{Y_{00}}^{-1}\left(F_{Y_{01}}(\cdot)\right)\right)}{f_{00}\left(F_{Y_{00}}^{-1}\left(F_{Y_{01}}(\cdot)\right)\right)}\check{\mathbb{G}}_{00}(\cdot).$$

Combining (D.9), (D.10), and the functional delta method, we have that

$$\sqrt{n}\left(DTT_n^{km}(\cdot)-DTT(\cdot)\right)\Rightarrow\mathbb{G}_{11}^{cic}(\cdot)-\tilde{\mathbb{G}}_{10}(\cdot)\ \text{ in } l^{\infty}\left(\mathcal{Y}_H\right).$$

The weak convergence of the quantile treatment effects on the treated follows from (D.9), (D.10), the fact the quantile mapping is Hadamard differentiable, and the functional delta method. More precisely, we have, under Condition 1[13],

(D.11)
$$\sqrt{n}\left(QTT_n^{km}(\cdot)-QTT(\cdot)\right)\Rightarrow\mathbb{G}_{11}^{q}(\cdot)-\tilde{\mathbb{G}}_{10}^{q}(\cdot)\ \text{ in } l^{\infty}((0,1)),$$

where

$$\begin{aligned}\tilde{\mathbb{G}}_{10}^{q}(\tau)&=\frac{\check{\mathbb{G}}_{00}^{q}(\tau)-\mathbb{G}_{01}^{cic}\left(F_{Y_{00}}\left(F_{Y_{10}}^{-1}(\tau)\right)\right)}{f_{01}\left(F_{Y_{00}}\left(F_{Y_{10}}^{-1}(\tau)\right)\right)},\\ \mathbb{G}_{11}^{q}(\tau)&=-\frac{\mathbb{G}_{11}^{cic}\left(F_{Y1}^{-1}(\tau)\right)}{f_{11}\left(F_{Y_{11}}^{-1}(\tau)\right)}\end{aligned}$$

and
$$\check{\mathbb{G}}_{00}^{q}(\tau)=\mathbb{G}_{00}^{cic}\left(F_{Y_{10}}^{-1}(\tau)\right)-\frac{f_{00}\left(F_{Y_{10}}^{-1}(\tau)\right)}{f_{10}\left(F_{Y_{10}}^{-1}(\tau)\right)}\mathbb{G}_{10}^{cic}\left(F_{Y_{10}}^{-1}(\tau)\right).$$

The asymptotic normality of the average treatment effect on the treated follows from (D.11), and the fact that

$$ATT=\int_0^1 QTT(u)\,du.$$

13 If Condition 1 is not satisfied, convergence is in $l^{\infty}((0,\bar{\tau}^{cic}))$

Thus

$$\sqrt{n}\left(ATT_n^{km} - ATT\right) \xrightarrow{d} \int_0^1 \left(\mathbb{G}_{11}^q(u) - \tilde{\mathbb{G}}_{10}^q(u)\right) du.$$

The validity of the bootstrap follows from Lo and Singh (1986), Stute (1995), and the functional delta method for the bootstrap, cf. Theorem 3.9.11 in van der Vaart and Wellner (1996). ■

E. Auxiliary Lemmas

In this section we present and prove some auxiliary lemmas that are useful for proving Propositions 2-4.

Lemma 1 *Let Assumption A.1 be satisfied. Then,* $\sup_x \left|p\left(x;\hat{\theta}_n\right) - p\left(x;\theta_0\right)\right| = o_{\mathbb{P}}\left(n^{-1/4}\right)$.

Proof of Lemma 1: Next, by classical mean value theorem, $p\left(x;\hat{\theta}_n\right) - p\left(x;\theta_0\right) = \dot{p}\left(x;\bar{\theta}\right)'\left(\hat{\theta}_n - \theta_0\right)$, where $\left|\left|\bar{\theta} - \theta_0\right|\right| \leq \left|\left|\hat{\theta}_n - \theta_0\right|\right|$ *a.s.*. Therefore,

$$\begin{aligned}
\left|p\left(x;\hat{\theta}_n\right) - p\left(x;\theta_0\right)\right|_\infty &= \left|\dot{p}\left(x;\bar{\theta}\right)'\left(\theta_n - \theta_0\right)\right|_\infty \\
&\leq \left|\dot{p}\left(x;\bar{\theta}\right)\right|_\infty \left|\left|\left(\theta_n - \theta_0\right)\right|\right| \\
&= o_{\mathbb{P}}\left(n^{-1/4}\right)
\end{aligned}$$

since $\left|\dot{p}\left(x;\bar{\theta}\right)\right|$ is uniformly bounded, and $\left|\left|\left(\hat{\theta}_n - \theta_0\right)\right|\right| = O_{\mathbb{P}}\left(n^{-1/2}\right)$. ■

Lemma 2 *Suppose Assumption 5.1 is satisfied. (i) If Assumption A.2 hold,* $\sup_{x\in\mathcal{X}}\left|\hat{p}_n^{\text{ker}}(x) - p_0(x)\right| = o_{\mathbb{P}}\left(n^{-1/4}\right)$. *(ii) If Assumption A.3 hold,* $\sup_{x\in\mathcal{X}}\left|\hat{p}_n^{\text{series}}(x) - p_0(x)\right| = o_{\mathbb{P}}\left(n^{-1/4}\right)$

Proof of Lemma 2: First, for the kernel estimator, it follows from Masry (1996b,a),

$$\begin{aligned}
\left|\hat{p}_n^{\text{ker}}(x) - p_0(x)\right|_\infty &= O_{\mathbb{P}}\left(b^s + \sqrt{\frac{\log n}{nb^k}}\right) \\
&= o_{\mathbb{P}}\left(b^{s/2}\right) \\
&= o_{\mathbb{P}}\left(n^{-1/4}\right),
\end{aligned}$$

where the second and this equalities follow from Assumptions A.2 (iii) and (iv).

For the Logit Series Estimator, define the pseudo true propensity score $p^{pseudo}(x) = L\left(R^L(x)'\pi_{0,L}\right)$

$$\pi_{0,L} = \arg\max_{\pi} \mathbb{E}\left[p_0(X)\log\left(\mathcal{L}\left(R^L(X)'\pi\right)\right) + (1-p_0(X))\log\left(1-\mathcal{L}\left(R^L(X)\pi\right)\right)\right].$$

Let $\zeta(L) = \sup_{x\in\mathcal{X}}\left|R^L(x)\right|$. Then, it follows from the triangle inequality, the mean value theorem, and Lemmas 1 and 2 of Hirano et al. (2003),

$$\begin{aligned}
\left|\hat{p}_n^{\text{series}}(x) - p_0(x)\right|_\infty &= \left|\hat{p}_n^{\text{series}}(x) - p^{pseudo}(x) + p^{pseudo}(x) - p_0(x)\right|_\infty \\
&\leq \left|\hat{p}_n^{\text{series}}(x) - p^{pseudo}(x)\right|_\infty + \left|p^{pseudo}(x) - p_0(x)\right|_\infty \\
&\leq c\zeta(L)\left\|\pi_{n,L} - \pi_{0,L}\right\| + \left|p^{pseudo}(x) - p_0(x)\right|_\infty \\
&= O_{\mathbb{P}}\left(\zeta(L)\left(\sqrt{\frac{L}{n}} + L^{-s/2r}\right)\right) \\
&= o_{\mathbb{P}}\left(n^{-1/4}\right)
\end{aligned} \tag{E.1}$$

where the last equality follow from Assumption A.3(iii). ■

Lemma 3 *The pathwise derivative of $S_1^{ate}(p_0)$, $S_0^{ate}(p_0)$, $S_1^{dte}(y,p)$ and $S_0^{dte}(y,p)$, with respect to p, $\Gamma_1^{ate}(p_0)[p-p_0]$, $\Gamma_0^{ate}(p_0)[p-p_0]$, $\Gamma_1^{dte}(y,p_0)[p-p_0]$, and $\Gamma_0^{dte}(p_0)[y,p-p_0]$, respectively, are given by*

$$\begin{aligned}
\Gamma_1^{ate}(p_0)[p-p_0] &= \mathbb{E}\left[-\frac{TY(p(X)-p_0(X))}{p_0^2(X)}\right], \\
\Gamma_0^{ate}(p_0)[p-p_0] &= \mathbb{E}\left[\frac{(1-T)Y(p(X)-p_0(X))}{(1-p_0(X))^2}\right], \\
\Gamma_1^{dte}(y,p_0)[p-p_0] &= -\mathbb{E}\left[\frac{T1\{Y\leq y\}(p(X)-p_0(X))}{p_0^2(X)}\right], \\
\Gamma_0^{dte}(y,p_0)[p-p_0] &= \mathbb{E}\left[\frac{(1-T)1\{Y\leq y\}(p(X)-p_0(X))}{(1-p_0(X))^2}\right],
\end{aligned}$$

in all directions $[p-p_0]\in\mathcal{H}$.

Proof of Lemma 3: The proof follows from noting that, for any $\bar{p}$ such that

$\{p_0 + \alpha(\bar{p} - p_0) : \alpha \in [0,1]\} \subset \mathcal{H}$,

$$\begin{aligned}
\frac{S_1^{ate}(p_0 + \alpha(\bar{p} - p_0)) - S_1^{ate}(p_0)}{\alpha} &= \frac{\mathbb{E}\left[\frac{TY}{p_0(X)+\alpha(\bar{p}(X)-p_0(X))} - \frac{TY}{p_0(X)}\right]}{\alpha} \\
&= \mathbb{E}\left[\frac{TY(\bar{p}(X) - p_0(X))}{p_0^2(X) + \alpha p_0(X)(\bar{p}(X) - p_0(X))}\right] \\
&\to \mathbb{E}\left[-\frac{TY(\bar{p}(X) - p_0(X))}{p_0^2(X)}\right] \text{ as } \alpha \to 0.
\end{aligned}$$

This establishes the proof of the representation of $\Gamma_1^{ate}(p_0)[p - p_0]$. The same arguments can be applied to establish the pathwise derivative of the other statistical functionals.■

Lemma 4 *Let Y_1 and Y_2 be square integrable random variables, and assume that $\varepsilon < p_0(\cdot) < 1 - \varepsilon$ a.s.. Then,*

$$\begin{aligned}
\left|S_1^{ate}(p_0) - S_1^{ate}(p_0) - \Gamma_1^{ate}(p_0)[p - p_0]\right| &\leq c \sup_{x \in \mathbb{X}} |p(x) - p_0(x)|^2, \\
\left|S_0^{ate}(p_0) - S_0^{ate}(p_0) - \Gamma_0^{ate}(p_0)[p - p_0]\right| &\leq c \sup_{x \in \mathbb{X}} |p(x) - p_0(x)|^2, \\
\sup_{y \in \mathbb{Y}} \left|S_1^{dte}(y, p_0) - S_1^{dte}(y, p_0) - \Gamma_1^{dte}(y, p_0)[p - p_0]\right| &\leq c \sup_{x \in \mathbb{X}} |p(x) - p_0(x)|^2, \\
\sup_{y \in \mathbb{Y}} \left|S_0^{dte}(y, p_0) - S_0^{dte}(y, p_0) - \Gamma_0^{dte}(y, p_0)[p - p_0]\right| &\leq c \sup_{x \in \mathbb{X}} |p(x) - p_0(x)|^2.
\end{aligned}$$

Proof of Lemma 4: Given that all functionals have symmetric construction, we only need to consider

$$\left|S_1^{ate}(p_0) - S_1^{ate}(p_0) - \Gamma_1^{ate}(p_0)[p - p_0]\right| \leq c \sup_{x \in \mathbb{X}} |p(x) - p_0(x)|^2.$$

To establish this inequality, not that

$$\begin{aligned}\left|S_1^{ate}(p_0) - S_1^{ate}(p_0) - \Gamma_1^{ate}(p_0)[p - p_0]\right| &= 2\left|\mathbb{E}\left[\frac{TY_1}{p_0^3(X)}(\bar{p}(X) - p_0(X))^2\right]\right| \\ &\leq 2\left|\mathbb{E}\left[\frac{TY_1}{p_0^3(X)}\right]\right|\sup_{x\in\mathbb{X}}|p(x) - p_0(x)|^2 \\ &\leq \frac{2}{\varepsilon^3}\mathbb{E}[|Y_1|]\sup_{x\in\mathbb{X}}|p(x) - p_0(x)|^2 \\ &\leq c\sup_{x\in\mathbb{X}}|p(x) - p_0(x)|\end{aligned}$$

where the first equality follows from a Taylor expansion argument, the first inequality follows from $|\bar{p}(X) - p_0(X)| \leq |\bar{p}(X) - p_0(X)|$, the second one from $p_0 > \varepsilon$ *a.s.*, and the fact that $T = \{0, 1\}$, and the last inequality from the square integrability of Y_1. ■

Lemma 5 *Let Assumptions 5.1 and A.1 be satisfied. Then, the classes of functions*

$$\begin{aligned}\mathcal{A}_{1,\psi}^{ate,p} &= \left\{y, x, t \to \psi_{1,p}^{ate} \equiv \frac{ty}{p(x;\theta)} : \theta \in \Theta\right\}, \\ \mathcal{A}_{0,\psi}^{ate,p} &= \left\{y, x, t \to \psi_{0,p}^{ate} \equiv \frac{(1-t)y}{1 - p(x;\theta)} : \theta \in \Theta\right\}, \\ \mathcal{A}_{1,\psi}^{dte,p} &= \left\{y, x, t \to \psi_{1,\bar{y},p}^{dte} \equiv \frac{t1\{y \leq \bar{y}\}}{p(x;\theta)} : \theta \in \Theta, \bar{y}\in\mathcal{Y}\right\}, \\ \mathcal{A}_{0,\psi}^{dte,p} &= \left\{y, x, t \to \psi_{0,\bar{y},p}^{dte} \equiv \frac{(1-t)1\{y \leq \bar{y}\}}{1 - p(x;\theta)} : \theta \in \Theta, \bar{y}\in\mathcal{Y}\right\},\end{aligned}$$

are Donsker.

Proof of Lemma 5: From the fact that for some $\varepsilon > 0$, $\varepsilon < p(\cdot) < 1-\varepsilon$ *a.s.*, the integrability conditions in Assumptions 5.1, and Theorem 2.10.6 of van der Vaart and Wellner (1996), it suffices to show that the classes of functions $\{1\{\cdot \leq \bar{y}\}, \bar{y}\in\mathcal{Y}\}, \{p(\cdot;\theta), \theta \in \Theta\}$, $\{y\}, \{t\}$ are Donsker. Since the functions $\{y\}$ and $\{t\}$ do not depend on θ nor on $\bar{y}$, they are clearly Donsker. From Theorem 19.5 of van der Vaart (1998) and Assumption A.1, the class of functions $\{p(\cdot;\theta), \theta \in \Theta\}$ is Donsker. The class of functions $\{1\{\cdot \leq \bar{y}\}, \bar{y}\in\mathcal{Y}\}$ is a VC-class, and hence it is Donsker, cf. Theorem 2.6.4 of van der Vaart and Wellner (1996). ■

Lemma 6 *Suppose Assumption 5.1 is satisfied, and let either Assumption A.2 or Assumption A.3 be true. Let $H = C_1^{\alpha}(\mathcal{X})$, where $C_1^{\alpha}(\mathcal{X})$ as in page 154 of van der Vaart and Wellner (2007). Then, the classes of functions*

$$\begin{aligned}
\mathcal{A}_{1,\psi}^{ate,p} &= \left\{ y,x,t \to \psi_{1,p}^{ate} \equiv \frac{ty}{p(x)} : p \in \mathcal{H} \right\}, \\
\mathcal{A}_{0,\psi}^{ate,p} &= \left\{ y,x,t \to \psi_{0,p}^{ate} \equiv \frac{(1-t)\,y}{1-p(x)} : p \in \mathcal{H} \right\}, \\
\mathcal{A}_{1,\psi}^{dte,p} &= \left\{ y,x,t \to \psi_{1,\bar{y},p}^{dte} \equiv \frac{t1\{y \le \bar{y}\}}{p(x)} : p \in \mathcal{H}, \bar{y} \in \mathcal{Y} \right\}, \\
\mathcal{A}_{0,\psi}^{dte,p} &= \left\{ y,x,t \to \psi_{0,\bar{y},p}^{dte} \equiv \frac{(1-t)\,1\{y \le \bar{y}\}}{1-p(x)} : p \in \mathcal{H}, \bar{y} \in \mathcal{Y} \right\},
\end{aligned}$$

are Donsker.

Proof of Lemma 6: Under the smoothness conditions imposed by Assumption A.2 or Assumption A.3, it follows that the class of functions $\{p, p \in \mathcal{H}\}$ is Donsker,cf. Theorem 2.7.1 of van der Vaart and Wellner (1996). Thus, following the same steps as in the proof of Lemma 5, we establish the desired results. ■

Lemma 7 *Let Assumptions 5.1 and A.1 be satisfied. Then,*

$$\begin{aligned}
\Gamma_1^{ate}(z,p_0)[p_n - p_0] &= \frac{1}{n}\sum_{i=1}^{n} \kappa_{1,i}^{ate,p} + o_{\mathbb{P}}\left(n^{-1/2}\right), \\
\Gamma_0^{ate}(z,p_0)[p_n - p_0] &= \frac{1}{n}\sum_{i=1}^{n} \kappa_{0,i}^{ate,p} + o_{\mathbb{P}}\left(n^{-1/2}\right)
\end{aligned}$$

where

$$\begin{aligned}
\kappa_{1,i}^{ate,p} &= \mathbb{E}\left[-\frac{m_1(X)}{p_0(X;\theta_0)} \dot{p}(X;\theta_0) \right]' I_{\theta_0}(T_i, X_i), \\
\kappa_{0,i}^{ate,p} &= \mathbb{E}\left[\frac{m_0(X)}{1-p_0(X;\theta_0)} \dot{p}(X;\theta_0) \right]' I_{\theta_0}(T_i, X_i),
\end{aligned}$$

$m_1(X) \equiv \mathbb{E}[Y_1|X]$, *and* $m_0(X) \equiv \mathbb{E}[Y_0|X]$. *Furthermore, the classes of functions*

$$\begin{aligned}
\mathcal{A}_1^{ate,p} &= \left\{ x,t \to \kappa_{1,}^{ate,p}(x,t;\theta) \equiv \mathbb{E}\left[-m_1(X)\, p_0(X;\theta_0)^{-1} \dot{p}(X;\theta_0) \right]' I_\theta(t,x) : \theta \in \Theta \right\}, \\
\mathcal{A}_0^{ate,p} &= \left\{ x,t \to \kappa_0^{ate,p}(x,t;\theta) \equiv \mathbb{E}\left[m_0(X)\,(1-p_0(X;\theta_0))^{-1} \dot{p}(X;\theta_0) \right]' I_\theta(t,x) : \theta \in \Theta \right\}
\end{aligned}$$

are Donsker.

Proof of Lemma 7: From Lemma 3, Assumption A.1, and Law of Iterated Expectations, we have that

$$\Gamma_1^{ate}(z,p_0)[p-p_0] = \mathbb{E}\left[-\frac{TY(p(X)-p_0(X))}{p_0^2(X)}\right] = \mathbb{E}\left[-\frac{m_1(X)(p(X;\theta)-p(X;\theta_0))}{p(X;\theta_0)}\right].$$

Next, because $E[-m_1(X)(p(X;\theta)-p(X;\theta_0))/p(X;\theta_0)]$ is linear in θ, it is Hadamard differentiable at θ_0 tangentially to Θ. Then, by the delta method, and Assumption A.1,

$$\begin{aligned}
\sqrt{n}\mathbb{E}\left[-\frac{m_1(X)(p(X;\theta_n)-p(X;\theta_0))}{p(X;\theta_0)}\right] &= \mathbb{E}\left[-\frac{m_1(X)\dot{p}(X;\theta_0)}{p(X;\theta_0)}\right]'\sqrt{n}(\theta_n-\theta_0)+o_{\mathbb{P}}(1)\\
&= \mathbb{E}\left[-\frac{m_1(X)\dot{p}(X;\theta_0)}{p(X;\theta_0)}\right]' n^{-1/2}\sum_{i=1}^{n} I_{\theta_0}(T_i,X_i)+o_{\mathbb{P}}(1),
\end{aligned}$$

concluding the proof of the linear representation of $\Gamma_1^{ate}(z,p_0)[p-p_0]$. Similar arguments apply to $\Gamma_0^{ate}(z,p_0)[p-p_0]$.

Given that $\kappa_1^{ate,p}(x,t,\theta)$ depends on θ only trough I_θ, and I_θ is Donsker by Assumption A.1, the Donsker property of $A_1^{ate,p}$ follows from the fact that $\mathbb{E}\left[-\frac{m_1(X)}{p_0(X;\theta_0)}\dot{p}(X;\theta_0)\right] \leq \mathbb{C} < \infty$ and Corollary 9.32 of Kosorok (2008). The result for $A_0^{ate,p}$ follows from the same arguments.

Lemma 8 *Let Assumptions 5.1 and A.1 be satisfied. Then,*

$$\begin{aligned}
\Gamma_1^{dte}(z,p_0)[p_n-p_0] &= \frac{1}{n}\sum_{i=1}^{n}\kappa_{1,i}^{dte,p}(y)+o_{\mathbb{P}}\left(n^{-1/2}\right),\\
\Gamma_0^{dte}(z,p_0)[p_n-p_0] &= \frac{1}{n}\sum_{i=1}^{n}\kappa_{0,i}^{dte,p}(y)+o_{\mathbb{P}}\left(n^{-1/2}\right),
\end{aligned}$$

where

$$\begin{aligned}
\kappa_{1,i}^{dte,p}(y) &= \mathbb{E}\left[-\frac{F_1(y|X)}{p_0(X;\theta_0)}\dot{p}(X;\theta_0)\right]' I_{\theta_0}(T_i,X_i),\\
\kappa_{0,i}^{dte,p}(y) &= \mathbb{E}\left[\frac{F_0(y|X)}{1-p_0(X;\theta_0)}\dot{p}(X;\theta_0)\right]' I_{\theta_0}(T_i,X_i),
\end{aligned}$$

$F_1(y|X) \equiv \mathbb{E}[1\{Y_1 \leq y\}|X]$, and $F_0(y|X) \equiv \mathbb{E}[1\{Y_0 \leq y\}|X]$. Furthermore, the classes of functions

$$\mathcal{A}_1^{dte,p} = \left\{x, t \to \kappa_1^{dte,p}(x, t; y, \theta) \equiv \mathbb{E}\left[-F_1(y|X)\, p_0(X;\theta_0)^{-1}\, \dot{p}(X;\theta_0)\right]' I_\theta(t, x) : y \in \mathcal{Y}, \theta \in \times\right\},$$
$$\mathcal{A}_0^{dte,p} = \left\{x, t \to \kappa_0^{dte,p}(x, t; y, \theta) \equiv \mathbb{E}\left[F_0(y|X)\,(1 - p_0(X;\theta_0))^{-1}\, \dot{p}(X;\theta_0)\right]' I_\theta(t, x) : y \in \mathcal{Y}, \theta \in \times\right\}$$

are Donsker.

Proof of Lemma 8: The proof for the linear representation of $\Gamma_1^{dte}(z, p_0)[p_n - p_0]$ and $\Gamma_0^{dte}(z, p_0)[p_n - p_0]$ follows the same steps of the proof of Lemma 7, and therefore is omitted.

Next, notice that $A_1^{dte,p}$ depends on θ only trough I_θ, which is Donsker by Assumption A.1, and on y only through $\mathbb{E}\left[-F_1(y|X)\, p_0(X;\theta_0)^{-1}\, \dot{p}(X;\theta_0)\right]$, which is a constant bounded function and therefore clearly Donsker. Thus, the Donsker property of $A_1^{dte,p}$ follows from Corollary 9.32 of Kosorok (2008). Similar arguments apply to $A_0^{dte,p}$.$\blacksquare$

Lemma 9 *Suppose Assumption 5.1 holds, and let either Assumption A.2 or Assumption A.3 be satisfied. Then,*

$$\sqrt{n}\Gamma_1^{ate}(z, p_0)[p_n - p_0] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \kappa_{1,i}^{ate,np} + o_p(1),$$
$$\sqrt{n}\Gamma_0^{ate}(z, p_0)[p_n - p_0] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \kappa_{0,i}^{ate,np} + o_{\mathbb{P}}(1),$$

where

$$\kappa_{1,i}^{ate,np} = -\frac{m_1(X_i)}{p_0(X_i)}(T_i - p_0(X_i)),$$
$$\kappa_{0,i}^{ate,np} = \frac{m_0(X_i)}{1 - p_0(X_i)}(T_i - p_0(X_i)),$$

$m_1(X) \equiv \mathbb{E}[Y_1|X]$, and $m_0(X) \equiv \mathbb{E}[Y_0|X]$. Furthermore, the classes of functions

$$\mathcal{A}_1^{ate,np} = \left\{x, t \to \kappa_1^{ate,np}(x, t; p) \equiv -\frac{m_1(x)}{p(x)}(t - p(x)) : p \in \mathcal{H}\right\},$$
$$\mathcal{A}_0^{ate,np} = \left\{x, t \to \kappa_0^{ate,np}(x, t) \equiv \frac{m_0(x)}{1 - p(x)}(t - p(x)) : p \in \mathcal{H}\right\}$$

are Donsker.

Proof of Lemma 9: From Lemma 3, and Law of Iterated Expectations, we have that

$$\Gamma_1^{ate}(z,p_0)[p-p_0] = \mathbb{E}\left[-\frac{TY(p(X)-p_0(X))}{p_0^2(X)}\right] = \mathbb{E}\left[-\frac{m_1(X)(p(X)-p_0(X))}{p_0(X)}\right].$$

From Ichimura and Linton (2005) we have that when one uses the Kernel estimator for the propensity score, we have

$$\begin{aligned}\sqrt{n}\mathbb{E}\left[-\frac{m_1(X)\left(p_n^{\ker}(X)-p_0(X)\right)}{p_0(X)}\right] &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_{1,i}^{ate,np} + O_{\mathbb{P}}\left(b^2 + b^{-1/2}n^{-1/2} + b^2 n^{1/2}\right)\\ &\quad + O_{\mathbb{P}}\left(b^4 n^{1/2} + b^{-1}n^{-1/2}\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_{1,i}^{ate,np} + o_{\mathbb{P}}(1),\end{aligned}$$

under the bandwidths conditions we have imposed in Assumption A.2.

Alternatively, when one uses the Logit Series estimator, using similar arguments as in Hirano et al. (2003)'s Addendum, we have

$$\begin{aligned}&\sqrt{n}\mathbb{E}\left[-\frac{m_1(X)\left(p_n^{\text{series}}(X)-p_0(X)\right)}{p_0(X)}\right]\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_{1,i}^{ate,np} + O_{\mathbb{P}}\left(\zeta(L)L^{-s/2k}\right) + O_{\mathbb{P}}\left(\frac{\zeta(L)^2}{\sqrt{n}}\right) + O_{\mathbb{P}}\left(\sqrt{n}\zeta(L)L_n^{-s/2k}\right)\\ &\quad + O_{\mathbb{P}}\left(\frac{\zeta(L)^{11/2}}{\sqrt{n}}\right) + O_{\mathbb{P}}\left(\max\left(L^{-1/2k}, \zeta(L_n)L^{-s/2k}\right)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_{1,i}^{ate,np} + O_{\mathbb{P}}\left(n^{-\left(\frac{s}{2k}-1\right)v}\right) + O_{\mathbb{P}}\left(n^{2v-\frac{1}{2}}\right) + O_{\mathbb{P}}\left(n^{-\left(\frac{s}{2k}-1\right)v+\frac{1}{2}}\right)\\ &\quad + O_{\mathbb{P}}\left(n^{\frac{11}{2}v-1/2}\right) + O_{\mathbb{P}}\left(\max\left(L^{1-\frac{s}{2k}}, L^{-\frac{1}{2k}}\right)\right)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \kappa_{1,i}^{ate,np} + o_{\mathbb{P}}(1),\end{aligned}$$

where the second equality follows from using power series and setting $L = a\cdot N^v$ as in Assumption A.3, and the last equality follows from the conditions we have imposed on v. Similar arguments apply to $\Gamma_1^{ate}(z, p_0)[p - p_0]$

The Donsker property of $A_1^{ate,np}$ follows from p being Donsker (see, Lemma 6), the fact that $\varepsilon < p(\cdot) < 1 - \varepsilon$ *a.s.* and Corollary 9.32 of Kosorok (2008). Similar arguments apply to $A_0^{ate,np}$.■

Lemma 10 *Suppose Assumption 5.1 is satisfied, and let either Assumption A.2 or Assumption A.3 be true. Then,*

$$
\begin{aligned}
\sqrt{n}\Gamma_1^{dte}(z, p_0)[p_n - p_0] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\kappa_{1,i}^{dte,np}(y) + o_{\mathbb{P}}(1), \\
\sqrt{n}\Gamma_0^{dte}(z, p_0)[p_n - p_0] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\kappa_{0,i}^{dte,np}(y) + o_{\mathbb{P}}(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa_{1,i}^{dte,np}(y) &= -\frac{F_1(y|X_i)}{1 - p_0(X_i)}(T_i - p_0(X_1)), \\
\kappa_{0,i}^{dte,np}(y) &= \frac{F_0(y|X_i)}{1 - p_0(X_i)}(T_i - p_0(X_1)),
\end{aligned}
$$

$F_1(y|x) \equiv \mathbb{E}[1\{Y_1 \leq y\}|X = x]$, *and* $F_0(y|x) \equiv \mathbb{E}[1\{Y_0 \leq y\}|X = x]$. *Furthermore, the classes of functions*

$$
\begin{aligned}
\mathcal{A}_1^{dte,np} &= \left\{x, t \mapsto \kappa_1^{dte,np}(x, t; y, p) \equiv -\frac{F_1(y|x)}{p(x)}(t - p(x)) : y \in \mathbb{R}, p \in \mathcal{H}\right\}, \\
\mathcal{A}_0^{dte,np} &= \left\{x, t \to \kappa_0^{dte,np}(x, t; y, p) \equiv \frac{F_0(y|x)}{1 - p(x)}(t - p(x)) : y \in \mathbb{R}, p \in \mathcal{H}\right\}
\end{aligned}
$$

are Donsker.

Proof of Lemma 10: The proof for the linear representation of $\Gamma_1^{dte}(z, p_0)[p_n - p_0]$ and $\Gamma_0^{dte}(z, p_0)[p_n - p_0]$ follows the same steps of the proof of Lemma 9, and therefore is omitted. The Donsker property of $A_1^{dte,np}$ and $A_0^{dte,np}$ follows from Corollary 9.32 of Kosorok (2008), Example 19.11 of van der Vaart (1998) and the fact that $\varepsilon < p(\cdot) < 1 - \varepsilon$ *a.s.*. ■ .

References

Beran, R. (1981): "Nonparametric regression with randomly censored survival data," *Technical Report.*

Breslow, N. E. (1972): "Discussion of the paper by D.R. Cox," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34, 216–217.

Callaway, B., T. Li, and T. Oka (2015): "Quantile Treatment Effects in Difference in Differences Models under Dependence Restrictions and with only Two Time Periods," *Mimeo.*

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71, 1591–1608.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2268.

Cox, D. (1975): "Partial Likelihood," *Biometrika*, 62, 269–276.

Cox, D. R. (1972): "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34, 187–220.

Dabrowska, D. M. (1989): "Uniform consistency of the kernel conditional Kaplan-Meier estimate," *The Annals of Statistics*, 17, 1157 – 1167.

Donald, S. G. and Y.-C. Hsu (2014): "Estimation and inference for distribution functions and quantile functions in treatment effect models," *Journal of Econometrics*, 178, 383–397.

Escanciano, J. C. (2006): "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Frandsen, B. R. (2015a): "Testing Censoring Point Independence," *Mimeo.*

——— (2015b): "Treatment Effects With Censoring and Endogeneity," *Journal of the American Statistical Association*, 110, 1745–1752.

García-Suaza, A. (2015): "Oaxaca-Blinder type Decomposition Methods for Duration Outcomes," *Mimeo.*

González-Manteiga, W. and C. Cadarso-Suárez (1994): "Asymptotic properties of a generalized kaplan-meier estimator with some applications," *Journal of Nonparametric Statistics*, 4, 65–78.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Heckman, J. J. and B. E. Honoré (1989): "The identifiability of the competing risks model," *Biometrika*, 76, 325–330.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

Ichimura, H. and O. Linton (2005): "Asymptotic expansions for some semiparametric program evaluation estimators," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, chap. 8, 149–170.

KAPLAN, E. L. AND P. MEIER (1958): "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 53, 457–481.

KOSOROK, M. R. (2008): *Introduction to empirical processes and semiparametric inference*, Springer.

LO, S.-H. AND K. SINGH (1986): "The product-limit estimator and the bootstrap: Some asymptotic representations," *Probability Theory and Related Fields*, 515, 481–515.

LOPEZ, O. (2011): "Nonparametric Estimation of the Multivariate Distribution Function in a Censored Regression Model with Applications," *Communications in Statistics - Theory and Methods*, 40, 2639–2660.

MASRY, E. (1996a): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 1, 571–599.

——— (1996b): "Multivariate regression estimation local polynomial fitting for time series," *Stochastic Processes and their Applications*, 65, 81–101.

PORTNOY, S. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.

SHORACK, G. R. AND J. A. WELLNER (1986): *Empirical processes with applications to statistics*, New York: Wiley.

STUTE, W. (1993): "Consistent estimation under random censorship when covariables are present," *Journal of Multivariate Analysis*, 45, 89 – 103.

——— (1995): "The central limit theorem under random censorship," *The Annals of Statistics*, 23, 422–439.

——— (1996): "Distributional convergence under random censorship when covariables are present," *Scandinavian Journal of Statistics*, 23, 461–471.

——— (1997): "Nonparametric model checks for regression," *The Annals of Statistics*, 25, 613–641.

STUTE, W., W. GONZÁLEZ-MANTEIGA, AND M. P. QUINDIMIL (1998): "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association*, 93, 141–149.

TSIATIS, A. A. (1975): "A nonidentifiability aspect of the problem of competing risks." *Proceedings of the National Academy of Sciences of the United States of America*, 72, 20–22.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*, New York: Springer.

——— (2007): "Empirical processes indexed by estimated functions," in *Asymptotics: Particles, Processes and Inverse Problems*, Beachwood, Ohio, USA: Institute of Mathematical Statistics, vol. 55, 234–252.