Doubly-Robust Quantile Treatment Effect Estimation

Benjamin I. Miller

October 14, 2022

Abstract

I develop a doubly-robust estimator of the quantile treatment effect on the treated (QTT). I modify the Callaway and Li (2019) conditional estimator of the QTT to obtain consistent estimates using either the propensity score or the conditional cdf of the first-differenced untreated outcomes. Aside from the benefits of obtaining consistent estimates of a QTT when a nuisance function is misspecified, there are also efficiency gains. In addition, assumptions on the smoothness of the nuisance parameters can be relaxed when the estimator is doubly-robust. I also show that asymptotically valid confidence intervals can be constructed using the empirical bootstrap. Then, I demonstrate via simulations that the Callaway and Li can produce a sharply higher root mean square error when compared to my estimator. Finally, I apply my estimator to estimate the effect of increasing the minimum wage on county-level unemployment rates, where I show significant and varied quantile treatment effects.

1 Introduction

There has been a recent reevaluation of the effectiveness of difference-in-difference estimators. Estimates of the average treatment effect on the treated (ATT) can be sensitive to nuisance functions that are based upon the probability of treatment given covariates, known as the propensity score, or the conditional mean of the difference in untreated outcomes before and after treatment for the untreated subpopulation. Identification of the ATT relies upon the parallel trends assumption and the overlap assumption. A relaxation of these assumptions involves conditioning on covariates when that may change either the expected treatment status or the change in untreated outcomes. These separate approaches can be combined to obtain a doubly-robust estimator of the ATT.

If we are willing to strengthen these assumptions, then we can go further than estimation of the ATT. The QTT can be estimated at any desired quantile. How useful this is depends upon the topic under study. If a researcher is concerned with the median outcome of a variable post-intervention, or outcomes at the tail of a variable's distribution, then an estimator of the QTT would be desired. Observing which quantiles experience a significant treatment effect could then alter policy responses. The assumptions that are necessary for identification go beyond the parallel trends and overlap assumptions. The parallel trends assumption is strengthened from an assumption on conditional mean independence to an assumption of independence conditioned on covariates. More assumptions are required on the outcome variables across time.

In this paper, I demonstrate that the strength of the assumptions from Callaway and Li (2019) purchase more than the authors may have realized. Using their assumptions, it is possible to generate a doubly-robust estimator of the QTT. This estimator then allows for a relaxation of the assumptions placed upon the nuisance functions. In this case, the nuisance functions are a propensity score and a conditional cdf of the difference in untreated outcomes from before and after treatment. The double-robustness property allows for a reduced rate of convergence of the both functions. If the nuisance functions are estimated nonparametrically, and depending upon the type of nonparametric estimator that is used, the double-robustness result has another beneficially property. Subject to minimal smoothness conditions on the nuisance functions, the rate of convergence itself may not matter

First, in this paper I provide the identification result that guarantees double-robustness. This result is an extension of the key identification result found in Callaway and Li (2019). The propensity score and the relevant conditional cdf can be misspecified, but not simultaneously. The properties of this estimator are studied. The properties are broken up into subcases, considering if the nuisance parameters are estimated nonparametrically or otherwise.

As a prerequisite to studying the properties of the QTT estimator, I derive the efficient influence function of the doubly-robust portion of my estimator. The portion that I am referring to is the cdf of the difference in untreated outcomes from the pre and post-treatment periods for the treated subpopulation at an arbitrary real number. It is upon the estimation of this parameter that the double-robustness result is applied. The efficient influence function is considered in the panel data setting.

With the efficient influence function in hand, I then demonstrate that the doubly-robust estimator achieves the semiparametric efficiency bound in the panel data setting. This is shown through two separate cases, where in the first case the nuisance functions are estimated parametrically. In the second case they are estimated nonparametrically, with the propensity estimated using the sieve logit estimator and the conditional cdf estimated using a kernel estimator. Both estimators could be chosen to be sieves or kernel estimators, and this would only change the restrictions on the smoothness of the nuisance functions.

Then, I demonstrate through simulations that without the double-robustness correction to the Callaway and Li (2019) estimator the root mean square error is very large, owing to the large standard error of the estimator. My doubly-robust estimator has root mean square errors that are less than one-third of the Callaway and Li (2019) estimator, even if the nuisance functions are misspecified.

I also show that when the empirical bootstrap is used for inference, the double-robustness property which allows for the weakening of the assumption that the nuisance functions converge to the truth at the rate of $o_p(n^{-1/4})$ no longer holds. For this reason, I maintain that the nuisance functions should converge at the rate of $o_p(n^{-1/4})$. I investigate when this assumption upon the rate of convergence holds for the relevant nuisance functions.

Finally, I apply my estimator to county-level unemployment data in order to compare it to the Callaway and Li (2019) estimator. I use the same dataset that is applied in Callaway and Li (2019). I then compare point estimates and confidence bounds between my estimator and the Callaway and

Li (2019) estimator.

1.1 Literature Review

This paper draws directly from Callaway and Li (2019) in order to establish the identification result and the form of the QTT estimator, but the idea of a doubly-robust estimator in this context was inspired by other papers in the treatment effects and missing data literature. Sant'Anna and Zhao (2020) examines the properties of the doubly-robust ATT estimator, and Callaway and Sant'Anna (2021) extends these results to the staggered treatment setting while also examining the asymptotic properties of various weighted ATT estimators. The doubly-robust ATT estimator has existed in the econometrics literature as an example of a doubly-robust estimator in Rothe and Firpo (2013), along with most of its properties in the single treatment period, panel data setting. The doubly-robust estimator combines the regression approach of Heckman, Ichimura, and Todd (1998) and Heckman et al. (1998), along with the propensity score matching approach of Abadie (2005), which itself is based upon Horvitz and Thompson (1952). My estimator is similar in that combines two approaches that are analogous to separate regression and propensity score approachesAll of these approaches take place within the difference-in-difference framework popularized by Card (1990) and Card and Krueger (1994).

The literature on quantile treatment effects, when considering either selection on observables or a panel data setting, prominently includes Firpo (2007), Athey and Imbens (2006), Bonhomme and Sauder (2011), and Chernozhukov et al. (2013). These approaches do not consider double-robustness, and their identification assumptions are stronger than the identification assumptions of this paper. For example, Firpo (2007) sets up an M-estimation problem with weights based upon propensity score matching, but identification depends upon a strong ignorability. assumption. The nonparametric logit sieve estimator of Hirano, Imbens, and Ridder (2003) is applied, but the conditions needed for asymptotic normality are considerably restrictive. A result similar to Firpo (2007) in the missing data setting is found in Wooldridge (2007).

The double-robustness properties of my estimator are based upon more than the aforementioned doubly-robust estimators of the ATT. A general weighting result for treatment effects is presented in Słoczyński and Wooldridge (2018), and the basis for constructing doubly-robust moments conditions is outlined in Chernozhukov et al. (2016). The latter work is closely related to the doubly-robust estimator in the missing data setting of Muris (2020). The construction of my double-robustness estimator is partly based upon the estimators in Sued, Valdora, and Yohai (2020) , though they consider a missing data setting and do not discuss the issue of inference. Rothe and Firpo (2019) and Rothe and Firpo (2013) consider the asymptotic properties of double-robust estimators when the nuisance functions are estimated using kernel density methods. The work of Fan et al. (2016) is particularly important for this paper, since it considers a doubly-robust estimators with nuisance parameters that are estimated via sieves. I also make one final point about a doubly-robust QTT estimator that exists in the literature. Caracciolo and Furno (2017) proposes an estimator of the quantile treatment effect (QTE) that involves taking a quantile of a random variable that is a function of the propensity score and fitted values; however, this estimator only identifies the quantile of interest in a very restrictive case, using unstated assumptions. Their approach partly builds off of Machado and Mata (2005), but this approach involves obtaining unconditional quantiles directly through a random sample over conditional quantiles.

1.2 Structure of the Paper

The paper is structured as follows. Section 2 lays presents the framework, assumptions and identification result. Section 3 presents the estimator and considers estimation of the nuisance parameters. Section 4 considers the large-sample properties of the estimator. Section 5 examines the validity of the empirical bootstrap when applying this estimator. Section 6 contains a Monte Carlo study that examines the small sample properties of my estimator at various quantiles under misspecification of the nuisance functions and in comparison to another esimator. Section 7 includes the application of my estimator to estimating quantile treatment effects on the treated using unemployment data. Section 8 concludes the paper. The mathematical proofs are contained in the appendix, as well as figures and tables.

2 Assumptions and Setting

2.1 Setting

The setting that I am considering is the panel data setting. As in Callaway and Li (2019), I assume that the data consists of at least three periods, with treatment period t and pre-treatment periods $t-1$ and $t-2$. No unit receives the binary treatment before time t. $D = 1$ for unit i if treated at time t. D=0 if an individual is never treated. The outcomes Y_t, Y_{t-1}, and Y_{t-2} are observed, along with covariates X.

Each unit i have the potential outcomes Y_{0t} and Y_{1t}, but these outcomes cannot be observed simultaneously for each unit i. The observed outcome Y_t is then expressed as

$$Y_t = DY_{1t} + (1-D)Y_{0t}$$

Untreated outcomes are observed in the previous periods, since treatment does not take place until period t. Then $Y_{t-1} = Y_{0t-1}$ and $Y_{t-2} = Y_{0t-2}$.

Let q_τ denote the τ-quantile for some random variable Z, where

$$q_\tau = F_Z^{-1}(\tau) := inf\{z : F_Z(z) \geq \tau\}$$

and $F_Z(z)$ is the cumulative distribution function (cdf) of Z. $F_{Y1t|D=1}$ and $F_{Y0t|D=1}$ denote the cdfs of the treated and untreated outcomes for the treated subpopulation, respectively. The QTT is then defined as,

$$QTT(\tau) = F^{-1}_{Y_{1t}|D=1}(\tau) - F^{-1}_{Y_{0t}|D=1}(\tau)$$

Interest in the QTT stems from the ability to identify the effect of an intervention on a treated group, compared to the counterfactual outcome. For example, suppose that a portion of a population receives a Covid-19 vaccine, with the outcome variable being gross income. While we may wish to know the treatment effect at the median ($\tau = 0.5$) for the entire population, the identification assumptions may too strong to identify such a parameter [1]. With weaker assumptions, we can identify QTT(0.5). That is to say, we can identify the median effect of Covid-19 vaccination on gross income for the treated subpopulation.

2.2 Identification Assumptions

These assumptions are directly from Callaway and Li (2019). When presenting these assumptions, I will note how they are comparatively mild when compared to other assumptions in the econometrics literature. Note that these assumptions are the distributional equivalent of the traditional diff-in-diff assumptions.

The first assumption is known as the "Copula Stability Assumption"

Assumption ID.1 (Copula Stability Assumption)**.** $C_{\Delta Y_{0t}, Y_{0t-1}|D=1,X}(\cdot,\cdot) = C_{\Delta Y_{0t-1}, Y_{0t-2}|D=1,X}(\cdot,\cdot)$

This assumption is the most controversial assumption used for identification. As explained in Callaway and Li (2019), this assumption requires both the panel data setting and data over three time periods. This assumption is not placing any restrictions on any of the marginal distributions of the variables involved. Instead, by placing restrictions on the copula we are restricting the joint distribution of the variables in question based upon their joint distribution in prior periods. We cannot observe the untreated outcome for the treated subpopulation, but we can jointly observe the outcomes in the periods before treatment for the treated subpopulation. By an application of Sklar's Theorem and by writing $Y_{0t}|D = 1$ as $Y_{0t} - Y_{0t-1} + Y_{0t-1}|D = 1$, the cdf of $Y_{0t}|D = 1$ can be identified using the joint distribution of $Y_{0t-1}|D = 1$ and $Y_{0t-2}|D = 1$.

This assumption is similar to, and perhaps even weaker than, the assumption of stationarity in the time-series setting. No claim is being made that a sequence of random variables has a constant joint distribution over some shift in time. All that is being claimed is that a feature of

[1] In the case of the average treatment effect vs the average treatment effect on the treated, see Wooldridge (2010)

the joint distribution, the joint dependence between the random variables, is fixed over a limited time period. A form of this assumption has been applied in the measurement error literature. In Cameron et al. (2004), the copula is used to model the difference in count variables, where each count variable represents different measurements of the same outcome.

The second assumption is directly exploited to obtain double-robustness of the estimator.

Assumption ID.2. $\Delta Y_{0t} \perp\!\!\!\perp D|X$

This assumption takes the parallel trends assumption of the difference-in-difference literature, $E[\Delta Y_{0t}|D = 1, X] = E[\Delta Y_{0t}|D = 0, X]$ and strengthens it over the entire distribution. The assumption states that the distribution of the untreated outcomes is unaffected by the treatment effect, conditional on the covariates. This assumption is necessary to obtain the estimator of $F_{\Delta Y_{0t}|D=1}(y)$. This assumption, as I will show, also makes the doubly-robust estimator of $F_{\Delta Y_{0t}|D=1}(y)$ the most efficient estimator in the panel data setting. Furthermore, as strong as this assumption is, it is weaker than the Strong Ignorability assumption of Rosenbaum and Rubin (1983), where $Y_{0t}, Y_{1t} \perp\!\!\!\perp D|X$, that is applied in Firpo (2007).

Now, let $\Delta Y_t = Y_t - Y_{t-1}$. Then, consider the following assumptions.

Assumption ID.3. Each of the random variables ΔY_t for the treated group and $\Delta Y_{t-1}, Y_{t-1}, Y_{t-2}$ for the treated group are continuously distributed on their support with densities that are uniformly bounded from above and bounded away from 0.

Assumption ID.4. The observed data $\{Y_{it}, Y_{it-1}, Y_{it-2}, X_i, D_i\}_{i=1}^n$ are independent and identically distributed draws from the joint distribution $F_{Y_t,Y_{t-1},Y_{t-2},X,D}$. In addition, $Y_{it} = D_i Y_{1it} + (1-D_i)Y_{0it}$, $Y_{it-1} = Y_{0it-1}$, $Y_{it-2} = Y_{0it-2}$.

Assumption ID.3 ensures the uniqueness of the copula by restricting the outcomes to be continuous. Assumption ID.4 restricts the data to be panel data. If the copula is not unique, then even with the Copula Stability Assumption we may not be able to identify the cdf of $Y_{0t}|D = 1$.

The next assumption is the final assumption that I use for identification. Let the support of X be denoted by $\mathcal{X}$.

Assumption ID.5. $p := P(D = 1) > 0$ and, for all $\mathbf{x} \in \mathcal{X}$, $p(\mathbf{x}) := P(D = 1|X = \mathbf{x}) < 1$.

The first part of the assumption ensures that there is some positive probability of treatment. The second assumption is the ”overlap” assumption that is common in the difference-in-difference literature. This guarantees that for any value of X in $\mathcal{X}$ there is a positive probability of that value appearing in both the control and treatment groups. Without this assumption, the QTT cannot be identified since $F_{\Delta Y_{0t}|D=1}(y)$ could not be identified for a population that contains that contains those values of X for which the assumption is violated. Note that the overlap assumption as stated here is not enough when the estimation of the propensity score is nonparametric. In that case, the propensity score requires sharp upper and lower bounds away from 0 and 1.

2.3 Identification

With the identification assumptions in hand, I now present the identification result.

Theorem 1. Under assumptions ID.1-ID.5, and assuming that $\pi(X) = p(X)$or $\tilde{P}(Y \leq y|X) = P(Y \leq y|X)$,

$$
\begin{aligned}
&F_{Y_{0t}|D=1}(y) \\
&= E\left[\mathbb{1}\{F^{-1}_{\Delta Y_{0t}|D=1}(F_{\Delta Y_{t-1}|D=1}(\Delta Y_{t-1})) \leq y - F^{-1}_{Y_{t-1}|D=1}(F_{Y_{t-2}|D=1}(Y_{t-2}))\}|D = 1\right]
\end{aligned}
$$

where,

$$
F_{\Delta Y_{0t}|D=1}(y) = E\left[\left(\frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)} - \frac{D}{p}\right)\tilde{P}(\Delta Y_{0t} \leq y|X)\right] \quad (1)
$$

if $\pi(X) = p(X)$ a.c., or $\tilde{P}(\Delta Y_{0t} \leq y|X) = P(\Delta Y_{0t} \leq y|X)$ a.c. Then,

$$
QTT(\tau) = F_{Y_{1t|D=1}}(\tau)^{-1} - F_{Y_{0t|D=1}}(\tau)^{-1}
$$

The proof of the first part of this result is provided in Callaway and Li (2019). For the sake of making this paper as self-contained as possible, I will outline their argument. First, note that since $E[\mathbb{1}_{Y_{0t|D=1}\leq y}] = E[\mathbb{1}_{Y_{0t|D=1}-Y_{0t-1|D=1}+Y_{0t-1|D=1}\leq y}] = E[\mathbb{1}_{\Delta Y_{0t|D=1}+Y_{0t-1|D=1}}]$. Since this expectation is over the joint distribution of the random variables $\Delta Y_{0t|D=1}$ and $Y_{0t-1|D=1}$, and since this joint distribution can be written in terms of the copula and the marginal distributions of Y_{0t-1} and Y_{0t-2} by assumption ID.1 and Sklar's Theorem. The result then follows from a change of variables.

The second part of the theorem is the basis for the double-robustness property of the estimator. Either the propensity score or the conditional cdf of ΔY_{0t} needs to be correctly specified so $F_{\Delta Y_{0t}|D=1}(\delta)$ will be correctly identified. $F_{\Delta Y_{0t}|D=1}(\delta)$ is needed to identify $F_{Y_{0t}|D=1}(y)$. The intuition behind the double-robustness result is that if the propensity score is correctly specified, then the information provided by conditional cdf of ΔY_{0t} becomes redundant, at least for identification. If $\tilde{P}(\Delta Y_{0t} \leq y|X)$ is correctly specified, then the weight that is applied to this conditional cdf filters out the incorrect information that is left over from misspecification of the propensity score.[2]

3 Estimation

This section will present the estimators that can be used to obtain the QTT under the identification assumptions. There are two different estimators that I present, with an asymptotically negligible difference. These estimators differ in that they calculate the weights differently for estimation of $F_{\Delta Y_{0t}|D=1}(y)$. The first estimator is,

$$Q\hat{T}T(\tau) = \hat{F}_{Y_{1t|D=1}}(\tau)^{-1} - \hat{F}_{Y_{0t|D=1}}(\tau)^{-1}$$

where

$$\hat{F}^{-1}_{Y_{1t}|D=1}(\tau) = inf\{y : \hat{F}_{Y_{1t}|D=1}(y) \geq \tau\}$$

[2] Unless otherwise noted, from this point onwards the nuisance functions will be assumed to be correctly specified from this point onward. So, $\pi(\mathbf{x}) = p(\mathbf{x})$and $\tilde{P}(Y \leq y|\mathbf{x}) = P(Y \leq y|\mathbf{x})$.

$$\hat{F}^{-1}_{Y_{0t}|D=1}(\tau) = inf\{y : \hat{F}_{Y_{0t}|D=1}(y) \geq \tau\}$$

$$\begin{aligned}&\hat{F}_{Y_{0t}|D=1}(y)\\&= n_{\mathcal{D}}^{-1}\sum_{i\in\mathcal{D}}[\mathbb{1}\{\hat{F}^{-1}_{\Delta Y_{0t}|D=1}(\hat{F}_{\Delta Y_{t-1}|D=1}(\Delta Y_{it-1})) \leq y - \hat{F}^{-1}_{Y_{t-1}|D=1}(\hat{F}_{Y_{t-2}|D=1}(Y_{it-2}))]\end{aligned}$$

where $n_{\mathcal{D}}$ denotes the number of treated observations and

$$\hat{F}_{\Delta Y_{0t}|D=1}(y) = n^{-1}\sum_{i=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n}D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_{it} \leq y\} - \left(\frac{1-D_i}{\frac{\sum_{k=1}^{n}D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n}D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i)\right] \tag{2}$$

An alternate estimator of $F_{\Delta Y_{0t}|D=1}(y)$ is,

$$\hat{F}_{\Delta Y_{0t}|D=1}(y) = n^{-1}\sum_{i=1}^{n}\left[\left(\frac{1-D_i}{l_0}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_{it} \leq y\} - \left(\frac{1-D_i}{l_0}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n}D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i)\right] \tag{3}$$

where $l_0 = n^{-1}\sum_{i=1}^{n}\frac{\pi(x_i)(1-D_i)}{1-\pi(x_i)}$. The interesting point here is the estimation of the nuisance functions. If the nuisance functions are estimated parametrically, then by standard assumptions in the appendix the nuisance functions will converge rapidly enough in probability to guarantee asymptotic normality, since the parameters that index the functions will converge at a sufficiently fast rate. The issue here is that it is unlikely for the nuisance functions to be correctly specified.

If the nuisance functions are estimated nonparametrically, then estimation depends upon exactly how they are estimated. For example, suppose that both nuisance functions are estimated using kernel-based methods. The advantage of this, as seen in Rothe and Firpo (2013), is that the kernel estimation permits the estimator to be decomposed into a bias term, a first-order stochastic term, and a second-order stochastic term. Depending upon how the bandwidth is chosen, the estimator can converge in probability at a fast enough rate to ensure asymptotic normality, but at a slower rate than would otherwise be necessary due to the double-robustness property.

The double-robustness property here does not only mean that the identifying moment is doubly-robust. An implication of this is that the higher-order derivatives are also doubly-robust.

This implies that like the identifying moment, the derivatives also equal 0 if at least one of the nuisance parameters is correctly specified. This is useful for both sieve and kernel estimation. In the case of kernel estimation, this implies that the bias term in the kernel decomposition equals 0. In the case of sieve estimation, the usefulness of this is that terms in the asymptotic expansion of the doubly-robust estimator can then be bounded by the bracket integral with respect to the L_2 norm over the function space.

4 Asymptotic Properties

The key asymptotic properties of the estimator revolve around the asymptotic behavior of the estimator of $F_{\Delta Y_{0t}|D=1}(y)$, in addition to the asymptotic behavior of $\pi(x)$ and $\hat{P}(\Delta Y_{0t} \leq y|X)$. The limiting behavior of the QTT estimator is unchanged by the doubly-robust estimator. Before the asymptotic behavior of the estimator can be discussed, the following assumption will be introduced:

Assumption NP.1.

$$\underset{\mathbf{x}\in\mathcal{X}}{sup}|\hat{\pi}(\mathbf{x}) - \pi(\mathbf{x})| = o_p(n^{-1/4})$$

$$\underset{\mathbf{x}\in\mathcal{X}}{sup}||\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}) - P(\Delta Y_{0t} \leq y|\mathbf{x})|| = o_p(n^{-1/4})$$

This is the conventional uniform convergence assumption in the literature on nonparametric rates of convergence. It is not necessary when the estimator is doubly robust, but it is sufficient. [3]. Parametric assumptions that imply Assumption NP.1 can be found in Appendix A. Nonparametric estimators for each of the nuisance functions, as well as the assumptions necessary to imply Assumption NP.1, can also be found in Appendix A.

Before I establish the asymptotic properties of the estimator, the efficient influence function needs to be found. Besides claiming that the estimator of $F_{\Delta Y_{0t}|D=1}(y)$ is doubly-robust, the efficient influence function will allow us to determine whether the estimator is the most efficient estimator

[3] This assumption is mentioned in a footnote of Callaway and Sant'Anna (2021) when the nuisance parameters are estimated nonparametrically, even though their estimator is doubly-robust. It is not necessary.

of $F_{\Delta Y_{0t}|D=1}(y)$. Note that this assumption of efficiency is only being made under the identification assumptions. If these assumptions do not hold, then the efficiency result fails. The efficient influence function should also be the identifying moment condition to estimate $F_{\Delta Y_{0t}|D=1}(y)$ at a fixed value y. When the nuisance functions are nonparametrically estimated, asymptotic normality will depend upon a Taylor expansion of the efficient influence functions.

The efficient influence functions is presented in the following theorem:

Theorem 2. Under assumptions ID.1-ID.5, the efficient influence function is,

$$
\begin{aligned}
F_\tau(W) =& \frac{(1-D)p(X)}{p(1-p(X))}\mathbb{1}_{\{\Delta Y_t \leq \delta\}} - \frac{(1-D)p(X)}{p(1-p(X))}P(\Delta Y_{0t} \leq \delta|X) \\
&+ \frac{D}{p}P(\Delta Y_{0t} \leq \delta|X) - \frac{D}{p}F_{\Delta Y_{0t}|D=1}(\delta)
\end{aligned}
$$

Note that if we were to take the expected value of this function, $\frac{E[D]}{p} = 1$, so in expectation the efficient influence function reduces to the identifying moment condition of $F_{\Delta Y_{0t}|D=1}(y)$ induced by $\hat{F}_{\Delta Y_{0t}|D=1}(y)$. With the efficient influence function in hand, we can proceed to describe the asymptotic behavior of $\hat{F}_{\Delta Y_{0t}|D=1}(y)$. Before that takes place, it should be noted how exactly each of the nuisance functions are estimated.

With the efficient influence function, I can now proceed to the first major distributional result. I will now proceed proving the consistency and asymptotic normality of $\hat{F}_{\Delta Y_{0t}|D=1}(y)$.

Theorem 3. Suppose that assumptions ID.1-ID.5 and NP.1 hold. Then $\hat{F}_{\Delta Y_{0t}|D=1}(y) \xrightarrow{p} F_{\Delta Y_{0t}|D=1}(y)$, and

$\sqrt{n}\left(\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(y)\right) \xrightarrow{d} N(0, E[\psi(D, p(X), P(\Delta Y_{0t} \leq y|X); p)]^2)$, where

$$
\psi(D, p(X), P(\Delta Y_{0t} \leq y|X); p) = \left[w_0(D_i, X_i; \gamma^*)(\mathbb{1}\{\Delta Y_{ti} \leq y\} - (w_0(D_i, X_i; \gamma^*) - w_1(D_i, X_i; \gamma^*))P(\Delta Y_{0ti} \leq y|X_i; \beta^*) - w_1(D_i)F_{\Delta Y_{0t}|D=1}(y)\right]
$$

and

$$
w_0(D, X; \gamma^*) = \left(\frac{1-D}{p}\frac{\pi(X; \gamma^*)}{1-\pi(X; \gamma^*)}\right)
$$

$$w_1(D) = \frac{D}{p}$$

Note that the result shows that the estimator attains the semiparametric efficient lower bound, since the asymptotic variance equals the second moment of the efficient influence function of $F_{\Delta Y_{0t}|D=1}(\delta)$.

Now, I will present a central limit theorem result, based upon a similar result in Callaway and Li (2019), which I will use to later establish the limiting behavior of the QTT estimator. Note the following result:

Proposition 1. Suppose assumptions ID.1-ID.5, and Assumption NP.1 holds. Then,

$$\left(\hat{G}_{\Delta Y_{0t}|D=1}, \hat{G}_{\Delta Y_{t-1}|D=1}, \hat{G}_{Y_{0t}|D=1}, \hat{G}_{Y_t|D=1}, \hat{G}_{Y_{t-1}|D=1}, \hat{G}_{Y_{t-2}|D=1}\right)$$
$$\xrightarrow{d} (\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)$$

In the space $S = l^\infty(\Delta\mathcal{Y}_{0t|D=1}) \times l^\infty(\Delta\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{0t|D=1}) \times l^\infty(\mathcal{Y}_{t|D=1}) \times l^\infty(\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{t-2|D=1})$ where $(\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)$ is a tight Gaussian process with mean 0 and covariance $V(y', y) = E[\eta(y)'\eta(y)]$ for $y = (y_1, y_2, y_3, y_4, y_5, y_6) \in S$ and with $\eta(y)$ given by

$$\eta(y) = \begin{bmatrix} \psi(D, p(X), P(\Delta Y_{0t} \leq y|X); p) \\ \frac{D}{p}\mathbb{1}\{\Delta Y_{t-1} \leq y_2\} - F_{\Delta Y_{t-1}|D=1}(y_2) \\ \frac{D}{p}\mathbb{1}\{\tilde{Y}_t \leq y_3\} - F_{Y_{0t}|D=1}(y_3) \\ \frac{D}{p}\mathbb{1}\{Y_t \leq y_4\} - F_{Y_t|D=1}(y_4) \\ \frac{D}{p}\mathbb{1}\{Y_{t-1} \leq y_5\} - F_{Y_{t-1}|D=1}(y_5) \\ \frac{D}{p}\mathbb{1}\{Y_{t-2} \leq y_6\} - F_{Y_{t-2}|D=1}(y_6) \end{bmatrix}$$

where

$$\hat{G}_{\Delta Y_{0t}|D=1}(\delta) = \sqrt{n}\left(\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(y)\right)$$
$$\tilde{Y}_{it} = F^{-1}_{\Delta Y_{0t}|D=1}\left(F_{\Delta Y_{t-1}|D=1}(\Delta Y_{it-1})\right) + F^{-1}_{\Delta Y_{t-1}|D=1}\left(F_{\Delta Y_{t-2}|D=1}(\Delta Y_{it-2})\right)$$

$$\tilde{F}_{Y_{0t}|D=1}(y) = \frac{1}{n_D}\sum_{i\in\mathcal{D}} \mathbb{1}\{\tilde{Y}_{it} \leq y\}$$

$$\tilde{G}_{Y_{0t}|D=1}(y) = \sqrt{n}\left(\tilde{F}_{Y_{0t}|D=1}(y) - F_{Y_{0t}|D=1}(y)\right)$$

Proposition SA2 and Theorem SA1 still hold from Callaway and Li (2019). I reproduce them here, in order to account for the change in notation and numbering of the assumptions, and the technical fact that the estimator of $F_{\Delta Y_{0t}|D=1}(\delta)$ has changed. Proposition 1 is used to establish the result in Proposition 2.

Proposition 2. Let $\hat{G}_0(y) = \sqrt{n}(\hat{F}_{Y_{0t}|D=1}(y)) - F_{Y_{0t}|D=1}(y))$ and let $\hat{G}_1(y) = \sqrt{n}(\hat{F}_{Y_t|D=1}(y)) - F_{Y_t|D=1}(y))$. Suppose assumptions ID.1-ID.5, and Assumption NP.1 hold. Then,

$$(\hat{G}_0, \hat{G}_1) \xrightarrow{d} (\mathbb{G}_0, \mathbb{G}_1)$$

where $\mathbb{G}_0$ and $\mathbb{G}_1$ are tight Gaussian processes with mean 0 with almost surely uniformly continuous paths on the space $\mathcal{Y}_{0t|D=1} \times \mathcal{Y}_{t|D=1}$ given by

$$\mathbb{G}_1 = \mathbb{V}_1$$

and

$$\mathbb{G}_0 = \mathbb{V}_0 + \int\left[\mathbb{W}_1 \circ K_2(y,v) - f_{\Delta Y_{0t}|D=1}\left(y - F^{-1}_{Y_{t-1}|D=1} \circ F_{Y_{t-2}|D=1} \circ \frac{\mathbb{W}_4 - \mathbb{W}_1 \circ K_1(v)}{f_{Y_{t-1}|D=1} \circ K_1(v)} - \mathbb{W}_2 \circ K_3(y,v)\right)\right]$$
$$\times \frac{f_{\Delta Y_{t-1}|Yt-2,D=1}(K_3(y,v)|v)}{f_{\Delta Y_{t-1}|D=1}(K_3(y,v))} dF_{Y_{t-2}|D=1}(v),$$

where $K_1(v) := F^{-1}_{Y_{t-1}|D=1} \circ F_{Y_{t-2}|D=1}(v)$, $K_2(y,v) := y - K_1(v)$,and $K_3(y,v) := F^{-1}_{\Delta Y_{t-1}|D=1} \circ F_{\Delta Y_{0t}|D=1}(K_2(y,v))$

The above proposition is then used to establish the limiting behavior of the QTT estimator.

Theorem 4. Suppose $F_{Y_{0t}|D=1}$ admits a positive continuous density $f_{Y_{0t}|D=1}$ on an interval $[a,b]$ containing an ϵ-enlargement of the set $\{F^{-1}_{Y_{0t}|D=1}(\tau) : \tau \in \mathcal{T}\}$. Suppose assumptions ID.1-ID.5, and

Assumption NP.1 hold. Then,

$$\sqrt{n}(\widehat{QTT}(\tau) - QTT(\tau)) \xrightarrow{d} \bar{\mathbb{G}}_1(\tau) - \bar{\mathbb{G}}_0(\tau)$$

where $(\bar{\mathbb{G}}_0(\tau), \bar{\mathbb{G}}_1(\tau))$ is a stochastic process in the metric space $(\ell^\infty(\mathcal{T}))^2$ with

$$\bar{\mathbb{G}}_0(\tau) = \frac{\mathbb{G}_0(F^{-1}_{Y_{0t}|D=1}(\tau))}{f_{Y_{0t}|D=1}(F^{-1}_{Y_{0t}|D=1}(\tau))} \qquad \bar{\mathbb{G}}_1(\tau) = \frac{\mathbb{G}_1(F^{-1}_{Y_t|D=1}(\tau))}{f_{Y_t|D=1}(F^{-1}_{Y_t|D=1}(\tau))}$$

The theorem is unchanged from Callaway and Li (2019), since the asymptotic distribution of $\hat{G}_0(y) = \sqrt{n}(\hat{F}_{Y_{0t}|D=1}(y)) - F_{Y_{0t}|D=1}(y))$ and $\hat{G}_1(y) = \sqrt{n}(\hat{F}_{Y_t|D=1}(y)) - F_{Y_t|D=1}(y))$ is unchanged by the doubly-robust estimator of $F_{\Delta Y_{0t}|D=1}(y)$. This is analogous to a doubly-robust estimator having the same distribution as other semiparametric two-step estimators [4].

5 The Bootstrap

The standard errors in the application are based upon the empirical bootstrap procedure. I assume that for the bootstrapped nuisance functions, denoted by *, the following assumption holds,

Assumption B.1.

$$\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\pi}^*(\mathbf{x}) - \hat{\pi}(\mathbf{x})| = o_{p^*}(n^{-1/4})$$

$$\sup_{\mathbf{x}\in\mathcal{X}} ||\hat{P}^*(\Delta Y_{0t} \leq \delta|\mathbf{x}) - \hat{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})|| = o_{p^*}(n^{-1/4})$$

This assumption gives the minimum rate of convergence of the difference between the bootstrapped estimator and the original estimator as they tend to zero. By an argument outlined in the appendix, the double-robustness property does not reduce the necessary rate of convergence to achieve asymptotic normality when applying the empirical bootstrap. The argument relies upon an expansion of the efficient influence function of the bootstrapped estimate of $F_{\Delta Y_{0t}|D=1}(y)$ for the

[4] See the introduction of Rothe and Firpo (2019)

bootstrapped sample around the original estimates of the nuisance functions. When the expansion using the full sample is around the true functions, the expected value of the pathwise derivatives with respect to the nuisance functions equal zero. This no longer holds at estimates of the true functions. This is analogous to the elimination of the bias term with kernel estimates of nuisance functions when considering the asymptotic behavior of double-robust estimators, as shown in Rothe and Firpo (2019). I then obtain the following proposition,

This result then establishes the following proposition,

Proposition 3. Suppose assumptions ID.1-ID.5, and Assumptions NP.1 and B.1 hold. Then,

$$\begin{aligned}&\left(\hat{G}^*_{\Delta Y_{0t}|D=1}, \hat{G}^*_{\Delta Y_{t-1}|D=1}, \hat{G}^*_{Y_{0t}|D=1}, \hat{G}^*_{Y_t|D=1}, \hat{G}^*_{Y_{t-1}|D=1}, \hat{G}^*_{Y_{t-2}|D=1}\right)\\ &\xrightarrow{d} (\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)\end{aligned}$$

In the space $S = l^\infty(\Delta\mathcal{Y}_{0t|D=1}) \times l^\infty(\Delta\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{0t|D=1}) \times l^\infty(\mathcal{Y}_{t|D=1}) \times l^\infty(\mathcal{Y}_{t-1|D=1}) \times l^\infty(\mathcal{Y}_{t-2|D=1})$ where $(\mathbb{W}_1, \mathbb{W}_2, \mathbb{V}_0, \mathbb{V}_1, \mathbb{W}_3, \mathbb{W}_4)$ is a tight Gaussian process with mean 0 and covariance $V(y', y) = E[\eta(y)'\eta(y)]$ for $y = (y_1, y_2, y_3, y_4, y_5, y_6) \in S$ and with $\eta(y)$ given by

$$\eta(y) = \begin{bmatrix} \psi(D, p(X), P(\Delta Y_{0t} \leq y|X); p \\ \frac{D}{p}\mathbb{1}\{\Delta Y_{t-1} \leq y_2\} - F_{\Delta Y_{t-1}|D=1}(y_2) \\ \frac{D}{p}\mathbb{1}\{\tilde{Y}_t \leq y_3\} - F_{Y_{0t}|D=1}(y_3) \\ \frac{D}{p}\mathbb{1}\{Y_t \leq y_4\} - F_{Y_t|D=1}(y_4) \\ \frac{D}{p}\mathbb{1}\{Y_{t-1} \leq y_5\} - F_{Y_{t-1}|D=1}(y_5) \\ \frac{D}{p}\mathbb{1}\{Y_{t-2} \leq y_6\} - F_{Y_{t-2}|D=1}(y_6) \end{bmatrix}$$

where * denotes the bootstrap analogue.

Using the above proposition, I then obtain the asymptotic behavior of the bootstrapped process.

Theorem 5. Under Assumptions ID.1-ID.5, B.1-B.2, and either Assumptions P.1-P.3, or Assumptions NP.1-NP.7 and C.1-C.4,

$$\sqrt{n}(\widehat{QTT}(\tau)^* - \widehat{QTT}(\tau)) \xrightarrow{d} \bar{\mathbb{G}}_1(\tau) - \bar{\mathbb{G}}_0(\tau)$$

where $(\bar{\mathbb{G}}_0(\tau), \bar{\mathbb{G}}_1(\tau))$ is a stochastic process in the metric space $(\ell^\infty(\mathcal{T}))^2$ with

$$\bar{\mathbb{G}}_0(\tau) = \frac{\mathbb{G}_0(F^{-1}_{Y_{0t}|D=1}(\tau))}{f_{Y_{0t}|D=1}(F^{-1}_{Y_{0t}|D=1}(\tau))} \qquad \bar{\mathbb{G}}_1(\tau) = \frac{\mathbb{G}_1(F^{-1}_{Y_t|D=1}(\tau))}{f_{Y_t|D=1}(F^{-1}_{Y_t|D=1}(\tau))}$$

where $(\mathbb{G}_0(\tau), \mathbb{G}_1(\tau))$ are as in Proposition 2.

6 Simulations

In this section I will present the estimation of the QTT at $\tau \in [0.2, 0.8]$, in increments of 0.02. The goal is to demonstrate not only how my estimator performs in small samples, but also how it compares to the estimator of Callaway and Li (2019). The data generating process is as follows: I generate the following data generating process with $N = 1000$ and $T = 3$ for 200 iterations.

$$v \sim Normal(0,1)$$

$$\eta|D = 0 \sim Norm(0,1) \qquad \eta|D = 1 \sim Normal(1,1)$$

$$X_1 \sim Uniform(0,1) \qquad X_2 \sim Uniform(-1,0)$$

$$X_3 \sim Uniform(-2,1) \qquad X_4 \sim Uniform(-1,0)$$

$$Y_{t-2} = 0.25X_1 + 0.5X_2 + 0.75X_3 + X_4 + \eta + v \qquad Y_{t-1} = 1 + 0.5X_1 + 0.75X_2 + X_3 + 1.5X_4 + \eta + v$$

$$Y_{0t} = 2 + 0.25X_1 + 0.5X_2 + 0.75X_3 + X_4 + \eta + v \qquad Y_{1t} = 1.5X_1 + X_2 + 1.5X_3 + X_4 + \eta + v$$

$$Y_t = D \times Y_{1t} + (1 - D) \times Y_{0t}$$

$$p(X,\gamma) = \frac{e^{\gamma X}}{1+e^{\gamma X}} \qquad \gamma = [-0.25, -0.5, -0.75, 1]$$

The data generating process is based upon Example 3 in Callaway and Li (2019). The distribution of the covariates is chosen so that, for the given values of the parameters, the probability of treatment and the conditional cdf of ΔY_{0t} do not output observed values that are too close to 0 and 1. This can cause numerical issues when inverting the estimators.

The parameters of the propensity score are estimated are estimated via mle. The parameters of the conditional cdf of ΔY_{0t}, β, are estimated via OLS regression of $\Delta Y_{0t}|D = 0$ on $\Delta X_t|D = 0$. σ, the standard deviation of Δv_t, is estimated by taking the standard deviation of the vector of residuals generated from the OLS regression.

The following estimator is applied,

$$\hat{F}_{\Delta Y_{0t}|D=1}(\delta) = \left[\sum_{i=1}^{n} \frac{p(\mathbf{x}_i,\hat{\gamma})(1-D_i)}{[1-p(\mathbf{x}_i,\hat{\gamma})]}\right]^{-1} \sum_{i=1}^{n} \frac{p(\mathbf{x}_i,\hat{\gamma})(1-D_i)}{[1-p(\mathbf{x}_i,\hat{\gamma})]}\left[\mathbb{1}\{Y_i \le \delta\} - \Phi\left(\frac{\delta-\hat{\beta}\mathbf{x}_i}{\hat{\sigma}}\right)\right] + n_{\mathcal{D}}^{-1}\sum_{i=1}^{n} D_i \Phi\left(\frac{\delta-\hat{\beta}\mathbf{x}_i}{\hat{\sigma}}\right)$$

where $\Phi(\cdot)$ is the standard normal cdf, and $\hat{\beta}, \hat{\gamma}$, and $\hat{\sigma}$ are the aforementioned estimates of the nuisance parameters. Note that this estimator normalizes the weights of the first term. This creates a normalized augmented inverse probability weighting estimator. This adds an asymptotically negligible normalization constant while improving the small sample performance of the estimator[5].

Here, misspecification of the propensity score is when the propensity score is chosen to be the standard normal cdf. Misspecification of the cdf nuisance function is considered when the function is chosen to be the logistic(0,1) cdf. In either case, each function resembles the true function over the support of the true underlying random variable that each function is based upon, but the misspecification is most pronounced in the tails. This misspecification is not far from the truth over what a researcher might observe based upon their data.

[5]For a discussion of the importance of normalization of inverse probability estimators, though not in the context of double-robustness, see Słoczyński, Uysal, and Wooldridge (2022). The main benefit of this normalization is a reduction in the small-sample bias of the estimator.

Table 1 in Appendix C compares the estimates of $QTT(0.5)$ across different specifications of the nuisance functions and estimators. $QTTdr$ represents the estimates when both nuisance functions are correctly specified. $QTTpro$ represents the estimates when the propensity score is correctly specified, but the conditional cdf is incorrectly specified. $QTTcdf$ represents the analogous estimates to $QTTpro$ when the conditional cdf is correctly specified. $QTTcl$ represents the Callaway and Li estimates. $QTTno$ represents when both nuisance functions are misspecified. Note that under mild misspecification of both nuisance functions the estimator has a similar performance to when only the propensity score is misspecified. The estimator is almost indistinguishable in performance compared to when only the propensity score is misspecified. Also, note that the standard error of the estimate using the Callaway and Li estimator is far greater than the standard error of the other estimates. This is a pattern that repeats across the quantiles.

The estimator that I propose in this paper outperforms the Callaway and Li estimator, at least under the data generating process that I used. Figure 1 shows that all variations of the estimators that I considered have a similar average absolute bias across the quantile estimates; however, as shown in Figure 2 in Appendix C, the root-mean-square error(RMSE) is comparable for the variations of my estimator under mild misspecification of the nuisance functions. The Callaway and Li estimator has an RMSE of approximately 1.48, regardless of the quantile.

7 Application

In this section, I use my method to study the effect of state-level changes in the minimum wage on county-level unemployment rates. The purpose of this application is to demonstrate how the standard errors of the estimates using my estimator compare to the standard errors of the estimates using the estimator of Callaway and Li (2019), and this application is based upon the application within that paper. Variations in state-level changes in minimum wage laws are exploited alongside variations in county-level observable characteristics, such as differences in population and median income. The goal is to examine the change in the distribution of county-level unemployment rates due to an increase in the minimum wage, and compare that to the distribution of unemployment rates had there been no change in the minimum wage.

The dataset that is used is taken from the replication materials of Callaway and Li (2019). They examine a period during which there was variation in state-level minimum wages, but the U.S. federal minimum wage remained flat until the end of the period. The outcome variable is the county-level unemployment rate, which they obtain from the Local Area Unemployment Statistics Database from the Bureau of Labor Statistics. County-level unemployment rates are available monthly, and they choose to use the February unemployment rates from 2005-2007, a month that they felt to be sufficiently far away from the federal wage change in July 2007. They merge county characteristics, the 1997 county median income and the 2000 county population, from the 2000 County Data Book [6]. The treatment group consists of counties in states, 11 states total excluding counties from states in the northeast, that increased their minimum wage by the first quarter of 2007. Counties in 20 states that did not increase their minimum wage by July 2007 form the control group.

The nuisance parameters are estimated parametrically. I assumed a logit specification for the propensity score, with the covariates chosen to be the natural log of county population, the natural log of the median county income, the squares of these terms, their interaction, factor variables for the South and West census regions, and the interaction of the factor variables with the other covariates. I assumed a probit specification for the conditional cdf of ΔY_{0t}, with the parameters estimated using ordinary least squares and assuming homoskedastic errors.

Due to the simultaneous estimation of parameters and construction of confidence intervals, I will construct uniform confidence bands as in Callaway and Li (2019). I outline the steps as an algorithm below:

1. For each $\tau \in \mathcal{T}$, calculate

$$\hat{\Sigma}(\tau)^{1/2} = (q_{0.75}(\tau) - q_{0.25}(\tau))/(z_{0.75} - z_{0.25})$$

[6] The replication materials can be found at https://onlinelibrary.wiley.com/doi/full/10.3982/QE935 .

This is equal to the bootstrap interquartile range divided by the interquartile range of a standard normal random variable, where $\hat{\Sigma}(\tau)$ is an estimate of the asymptotic variance of $\widehat{QTT}(\tau)$.

2. For bootstrap iterations $b = 1, ..., B$, calculate,

$$I^b = sup_{\tau \in \mathcal{T}} \hat{\Sigma}(\tau)^{-1/2} | \sqrt{n}(\widehat{QTT}(\tau)^b - \widehat{QTT}(\tau))|$$

3. Calculate $c^B_{1-\alpha}$, which is the $(1-\alpha)$ quantile of $\{I^b\}^B_{b=1}$.

4. Calculate $\widehat{QTT}(\tau) \pm c^B_{1-\alpha} \hat{\Sigma}(\tau)^{1/2} / \sqrt{n}$

Figure 3 shows that when comparing the estimates, there is little difference between my estimator and the Callaway and Li (2019) estimator. The point estimators are close, except at the 90th quantile, and the confidence intervals based upon my estimator are only slightly narrower. This is more revealing than it may seem. My simulations would suggest a sharp reduction in the standard error of the estimates when applying my estimator, but this is only when the nuisance function estimates are sufficiently close to the truth. My estimator suggests that there is a strong misspecification of the conditional cdf of ΔY_{0t}.

8 Conclusion

I have provided a doubly-robust estimator of the quantile treatment effect on the treated. This estimator relaxes the assumptions on the nuisance functions, allowing for a slower rate of convergence in order to achieve the limiting distribution of the QTT estimator. This causes nonparametric estimation of each of the nuisance functions to be much more viable, since nonparametric estimation requires assumptions upon the differentiable of the nuisance functions, which in turn affects the rate of convergence. As my simulations demonstrate, this leads to a lower RMSE in small samples, particularly when estimating the QTT at the median. Without the double-robustness property, confidence intervals could be so large that the QTT is not statistically different from 0 except at the extremes of the distribution of the difference in treated and untreated outcomes for the treated subpopulation.

It is important to recognize what this estimator is not. It is not a substitute for the doubly-robust estimator of the ATT that is presented in Sant'Anna and Zhao (2020). The assumptions necessary for identification are relaxed. There is no conditional copula assumption, and the parallel trends assumption is weaker than the conditional independence assumption on the difference in untreated outcomes. Instead, the two estimators should be used to complement each other. The ATT should be estimated along with quantile treatment effects on the treated at a variety of quantiles. If the estimate of the ATT is inconsistent with the results that are being presented across the information that is summarized by the QTTs, then perhaps either the conditional copula assumption or the conditional independence assumption does not hold.

What this estimator should be seen as is part of a middle ground between some of the more nonparametric estimators and estimators that rely entirely upon propensity score matching. In particular, the optimal transport methods of Gunsilius and Xu (2021) and Torous, Gunsilius, and Rigollet (2021) avoid the curse of dimensionality that is common with nonparametric estimation of the propensity score when estimating treatment effects; however, a doubly-robust estimator will relax the smoothness assumptions on the propensity score function in relation to the dimension of the covariate matrix. When supplemented with other doubly-robust estimators in the causal inference literature, my QTT estimator becomes part of a battery of doubly-robust estimators that increase the feasibility of propensity score matching.

References

Abadie, Alberto (2005). "Semiparametric difference-in-differences estimators". In: *The Review of Economic Studies* 72.1, pp. 1–19.

Athey, Susan and Guido W Imbens (2006). "Identification and inference in nonlinear difference-in-differences models". In: *Econometrica* 74.2, pp. 431–497.

Bonhomme, Stéphane and Ulrich Sauder (2011). "Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling". In: *Review of Economics and Statistics* 93.2, pp. 479–494.

Callaway, Brantly and Tong Li (2019). "Quantile treatment effects in difference in differences models with panel data". In: *Quantitative Economics* 10.4, pp. 1579–1618.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Cameron, A Colin et al. (2004). "Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts". In: *The Econometrics Journal* 7.2, pp. 566–584.

Caracciolo, Francesco and Marilena Furno (2017). "Quantile treatment effect and double robust estimators: an appraisal on the Italian labor market". In: *Journal of Economic Studies*.

Card, David (1990). "The impact of the Mariel boatlift on the Miami labor market". In: *ILR Review* 43.2, pp. 245–257.

Card, David and Alan Krueger (1994). *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*.

Chen, Xiaohong (2007). "Large sample sieve estimation of semi-nonparametric models". In: *Handbook of econometrics* 6, pp. 5549–5632.

Chen, Xiaohong, Oliver Linton, and Ingrid Van Keilegom (2003). "Estimation of semiparametric models when the criterion function is not smooth". In: *Econometrica* 71.5, pp. 1591–1608.

Chen, Xiaohong and Xiaotong Shen (1998). "Sieve extremum estimates for weakly dependent data". In: *Econometrica*, pp. 289–314.

Chernozhukov, Victor et al. (2013). “Average and quantile effects in nonseparable panel models”. In: *Econometrica* 81.2, pp. 535–580.

Chernozhukov, Victor et al. (2016). “Locally robust semiparametric estimation”. In: *arXiv preprint arXiv:1608.00033*.

Fan, Jianqing et al. (2016). *Improving covariate balancing propensity score: A doubly robust and efficient approach*. Tech. rep. Technical report, Princeton University.

Firpo, Sergio (2007). “Efficient semiparametric estimation of quantile treatment effects”. In: *Econometrica* 75.1, pp. 259–276.

Gunsilius, Florian and Yuliang Xu (2021). “Matching for causal effects via multimarginal optimal transport”. In: *arXiv preprint arXiv:2112.04398*.

Heckman, James J, Hidehiko Ichimura, and Petra Todd (1998). “Matching as an econometric evaluation estimator”. In: *The review of economic studies* 65.2, pp. 261–294.

Heckman, James J et al. (1998). *Characterizing selection bias using experimental data*.

Hirano, Keisuke, Guido W Imbens, and Geert Ridder (2003). “Efficient estimation of average treatment effects using the estimated propensity score”. In: *Econometrica* 71.4, pp. 1161–1189.

Horvitz, Daniel G and Donovan J Thompson (1952). “A generalization of sampling without replacement from a finite universe”. In: *Journal of the American statistical Association* 47.260, pp. 663–685.

Li, Qi and Jeffrey S Racine (2008). “Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data”. In: *Journal of Business & Economic Statistics* 26.4, pp. 423–434.

Li, Qi and Jeffrey Scott Racine (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Lorentz, GG (1966). *Approximation of Functions, Athena Series*. Holt, Rinehart and Winston, New York.

Machado, José AF and José Mata (2005). “Counterfactual decomposition of changes in wage distributions using quantile regression”. In: *Journal of applied Econometrics* 20.4, pp. 445–465.

Masry, Elias (1996). “Multivariate local polynomial regression for time series: uniform strong consistency and rates”. In: *Journal of Time Series Analysis* 17.6, pp. 571–599.

Muris, Chris (2020). “Efficient GMM estimation with incomplete data”. In: *Review of Economics and Statistics* 102.3, pp. 518–530.

Newey, Whitney K (1990). “Semiparametric efficiency bounds”. In: *Journal of applied econometrics* 5.2, pp. 99–135.

Rosenbaum, Paul R and Donald B Rubin (1983). “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1, pp. 41–55.

Rothe, Christoph and Sergio Firpo (2013). “Semiparametric estimation and inference using doubly robust moment conditions”. In:

Rothe, Christoph and Sergio Firpo (2019). “Properties of doubly robust estimators when nuisance functions are estimated nonparametrically”. In: *Econometric Theory* 35.5, pp. 1048–1087.

Sant’Anna, Pedro HC and Jun Zhao (2020). “Doubly robust difference-in-differences estimators”. In: *Journal of Econometrics* 219.1, pp. 101–122.

Słoczyński, Tymon, S Derya Uysal, and Jeffrey M Wooldridge (2022). “Abadie’s Kappa and Weighting Estimators of the Local Average Treatment Effect”. In: *arXiv preprint arXiv:2204.07672*.

Słoczyński, Tymon and Jeffrey M Wooldridge (2018). "A general double robustness result for estimating average treatment effects". In: *Econometric Theory* 34.1, pp. 112–133.

Sued, Mariela, Marina Valdora, and Víctor Yohai (2020). "Robust doubly protected estimators for quantiles with missing data". In: *TEST* 29.3, pp. 819–843.

Torous, William, Florian Gunsilius, and Philippe Rigollet (2021). "An Optimal Transport Approach to Causal Inference". In: *arXiv preprint arXiv:2108.05858*.

Vaart, Aad W and Jon A Wellner (1996). "Weak convergence". In: *Weak convergence and empirical processes*. Springer, pp. 16–28.

Vaart, Aad W Van der (2000). *Asymptotic statistics*. Vol. 3. Cambridge university press.

Wooldridge, Jeffrey M (2007). "Inverse probability weighted estimation for general missing data problems". In: *Journal of econometrics* 141.2, pp. 1281–1301.

— (2010). *Econometric analysis of cross section and panel data*. MIT press.

Appendices

A High-Level Assumptions and Propositions for Nuisance Function Estimation

Assumptions P.1-P.3 are the parametric assumptions that are sufficient to imply Assumption NP.1.

Assumption P.1. (i)$G(x;\gamma)$ is a parametric model for $p(x)$, where $\gamma \in \Gamma \subset \mathbb{R}^M$ and $G(x,\gamma) > 0$, all $x \in X, \gamma \in \Gamma$, where Γ is compact. (ii) There exists $\gamma_0 \in \Gamma$ such that $p(x) = G(x,\gamma_0), \gamma_0 \in$ $int(\Gamma)$. (iii)$G(X;\gamma)$ is a.s. twice continuously differentiable in a neighborhood of $\gamma_0, \Gamma^* \subset \Gamma$. (iv)$\hat{\gamma}$ is a consistent estimator of γ_0 and $n^{1/2}(\hat{\gamma} - \gamma_0) = n^{-1/2}\sum_{i=1}^{n} l_\gamma(W_i;\gamma_0) + o_p(1)$, where $W_i =$

$(Y_{01}, Y_{i1}, D_i, X_i)$, $E[l_\gamma(W_i;\gamma_0)] = 0, E[l_\gamma(W_i;\gamma_0)l_\gamma(W_i;\gamma_0)'] $ exists and is positive definite and $\lim_{\delta\to 0}$ $\mathrm{E}[\sup_{\gamma\in\Gamma^*,||\gamma-\gamma^*||\leq\delta}||l_\gamma(W_i;\gamma) - l_\gamma(W_i;\gamma_0)||^2] = 0$. (vi) For some $\epsilon > 0, 0 < P(x;\gamma) \leq 1-\epsilon$ a.s. for all $\gamma \in int(\Gamma)$.

Assumption P.2. (i)$g(x) = g(x;\beta)$ is a parametric model for the conditional mean of Y_{0t}, where $\beta \in \Theta \subset \mathbb{R}^k, \Theta$ being compact; (ii) $g(X,\beta)$ is a.c. continuous at each $\beta \in \Theta$; (iii) there exists a unique pseudo-true parameter $\beta^* \in int(\Theta)$;(iv)$g(X,\beta)$ is a.c. twice continuously differentiable in a neighborhood of $\beta^*, \Theta^* \subset \Theta$; (iv) the estimator $\hat{\beta}$ is strongly consistent for β^* and satisfies the following linear expansion:

$$\sqrt{n}(\hat{\beta}-\beta^*) = n^{-1/2}\sum_{i=1}^{n} l_\beta(W_i;\beta^*) + o_p(1)$$

where $l_\beta(\cdot;\beta)$ is such that $E[l_\beta(W;\beta^*)] = 0, E[l_\beta(W;\beta^*)l_g(W;\beta^*)']$ exists and is positive definite and $lim_{a\to 0}E\left[sup_{\beta\in\Theta^*,||\beta-\beta^*\leq a}||l_\beta(W;\beta) - l_\beta(W;\beta^*)||\right] = 0$.

Assumption P.3. $E[||h(W;\beta,\gamma)||^2] < \infty$ and $E\left[sup_{\beta\in\Theta_s,\gamma\in\Gamma_s}|\dot{h}(W;\beta,\gamma)\right] < \infty$ where Θ^s, Γ_s is a small neighborhood of β^*, γ^*, and

$$h(W;\beta,\gamma) = (w_0(D,X;\gamma))\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - (w_0(D,X;\gamma) - w_1(D))P(\Delta Y_{0t} \leq \delta, X;\beta)$$

These are the standard assumptions found in the literature, such as in Sant'Anna and Zhao (2020). Assumptions P.2 and P.2 imply that the parameters which index $p(x;\gamma)$ and $P(\Delta Y_{0t} \leq \delta|x;\beta)$ are sufficiently smooth and are $\sqrt{n}$-asymptotically linear. Assumption P.3 is an integrability condition. Assumptions P.1 and P.2 are stronger than Assumption NP, while Assumption P.3 is necessary to apply the Weak Law of Large Numbers along with the Central Limit Theorem.

I consider as a nonparametric estimator of the propensity score the sieve logit estimator of Hirano, Imbens, and Ridder (2003), though the proof and assumptions that I am placing on that estimator are different, and in some sense relaxed, compared to the conditions in Hirano, Imbens, and Ridder (2003) that are used to prove that the estimator converges uniformly to the true function at

$o_p(n^{-1/4})$. When using this estimator, the goal is to approximate $p(\mathbf{x})$,using a series approximation such that

$$m(\mathbf{x})_0 \approx \tilde{r}^{\tilde{K}}(\mathbf{x})'\Gamma_{\tilde{K}}$$

where

$$\tilde{r}^{\tilde{K}}(\mathbf{x})' = (r_{1\tilde{K}}(\mathbf{x}), ..., r_{\tilde{K}\tilde{K}}(\mathbf{x}))'$$
$$K = \tilde{K} + 1$$

The estimator is given by

$$m^* = argmax_{\pi \in \mathcal{H}_n} \frac{1}{n} \sum_{i=1}^{n} [D_i logL(m(\mathbf{x}_i)) + (1 - D_i) log(1 - L(m(\mathbf{x}_i))]$$

where $L(a) = \frac{exp(a)}{1+exp(a)}$ and $\mathcal{H}_n$ denotes the sieve space. Let the sieve space be over the s-smooth class of functions which I will denote by,

$$\mathcal{H} = \Lambda_c^p(\mathcal{X}) = \left\{ m \in C^s(\mathcal{X}) : \sup_{[\alpha] \le s} \sup_{x \in \mathcal{X}} |D^\alpha m(x)| \le c, \sup_{[\alpha]=s} \sup_{x,y \in \mathcal{X}, x \ne y} \frac{D^\alpha m(x) - D^\alpha h(y)}{|x-y|_e^\gamma} \le c \right\}$$

where $C^s(\mathcal{X})$ denotes the space of all s-times continuously differentiable functions on $\mathcal{X}$, and $|\cdot|$ denotes the Euclidean norm. Furthermore, let $\mathcal{H}_n = \left\{ m \in \mathcal{H}_n : m(\mathbf{x}) = \tilde{r}^{\tilde{K}}(\mathbf{x})'\Gamma_{\tilde{K}}, |m|_s \le c \right\}$.

I will let $||m - m_0||_\infty = \sup_{\mathbf{x} \in \mathcal{X}} |m(\mathbf{x}) - m_0(\mathbf{x})|$ and $\ell(m, \mathbf{x}_i) = D_i logL(m(\mathbf{x}_i)) + (1 - D_i) log(1 - L(m(\mathbf{x}_i))$. Let $H(w, \mathcal{F}_n, ||\cdot||_r)) := log(N(w, \mathcal{F}_n, ||\cdot||_r))$, where $N(w, \mathcal{F}_n, ||\cdot||_r)$ is the minimal number of w-balls that cover $\mathcal{F}_n$ under $||\cdot||_r$, and

$\mathcal{F}_n = \{\ell(m, \mathbf{x}_i) - \ell(m, \mathbf{x}_i) : ||m - m_0|| \le \delta, m \in \mathcal{H}_n\}$. In addition,

$\delta_n = \inf\left\{ \delta \in (0,1) : \frac{1}{\sqrt{n}\delta^2} \int_{b\delta^2}^{\delta} \sqrt{H(w, \mathcal{F}_n, ||\cdot||_r))} dw \le const. \right\}$, where $b > 0$ is a constant.

The assumptions below are sufficient for the consistency of the sieve logit estimator and to satisfy Assumption NP.1. They are based upon conditions in Chen (2007):

Assumption NP.2. (i)$E[D(logL(m_0(\mathbf{x}))) + (1 - D)log(1 - L(m_0(\mathbf{x}))] > -\infty$, and

if $E[D(logL(m_0(\mathbf{x}))) + (1 - D)log(1 - L(m_0(\mathbf{x}))] = \infty$

then $E[D(logL(m_0(\mathbf{x}))) + (1 - D)log(1 - L(m_0(\mathbf{x}))] < \infty$ for all $m \in \mathcal{H}_k \backslash m_0$ for all $k \geq 1$

(ii)There are functions d() and t(), where d() is a non-increasing positive function and t() is a positive function such that for all $\epsilon > 0$ and for all $k \geq 1$,

$E[D(logL(m_0(\mathbf{x}))) + (1 - D)log(1 - L(m_0(\mathbf{x}))]$

- $sup_{m \in \mathcal{H}_n : ||m - m_0||_\infty \geq \epsilon} E[D(logL(m(\mathbf{x}))) + (1 - D)log(1 - L(m(\mathbf{x}))]$

$\geq d(k)t(\epsilon) > 0$

Assumption NP.3. $\mathcal{H}_k \subset \mathcal{H}_{k+1} \subset \mathcal{H}$ for all $k \geq 1$, and there exists a sequence $\sigma_k m_0 \in \mathcal{H}_k$ such that $||\sigma_k m_0 - m_0||_\infty \rightarrow 0$ as $k \rightarrow \infty$.

Assumption NP.4. (i)The sieve spaces $\mathcal{H}_k$ are compact under $||m_1 - m_2||_\infty$, where $m_1, m_2 \in \mathcal{H}_k$ (ii)lim $inf_{k(n)} d(k(n)) > 0$, $E[\ell(h, \mathbf{x}_i]$ is continuous at $m = m_0 \in \Pi$, and $E[sup_{m \in \mathcal{H}_n} |\ell(m, \mathbf{x}_i)]$ is bounded. (iii)$E[||\mathbf{x}_i||] < \infty$

Assumption NP.5. $log(N(\delta, \mathcal{H}_n, || \cdot ||)) = o(n)$ for all $\delta > 0$.

Assumption NP.6. There exist $\underline{p}$ and $\overline{p}$ such that $0 < \underline{p} \leq p(\mathbf{x}) \leq \overline{p} < 1$.

Assumption NP.7. $\frac{2(a)^2}{(2a+d)^2} > \frac{1}{4}$, where d is the dimension of X, and $a = (s + \alpha)$, where $m_0(\mathbf{x})$ is a times continuously differentiable and $|m_0(\mathbf{x}) - m_0(\mathbf{y})| \leq ||\mathbf{x} - \mathbf{y}||_e^\alpha, \mathbf{x}, \mathbf{y} \in \mathcal{X}$ under the Euclidean norm $|| \cdot ||_e$ for $0 < \alpha \leq 1$.

Assumptions NP.2 consists of regularity conditions on the objective function. Assumption NP.3 implies that there exists some sequence of functions such that on subsets of the entire function space there exists some sequence of function that uniformly converge to 0 as the subspaces grow in size. Assumption NP.4 implies the existence of a solution at which the objective function is minimized. Assumption NP.5 ensures that the function space does not grow too fast as the sample size increases. Assumption NP.6 strengthens Assumption ID.5 so that the propensity score has upper and lower bounds away from 0 and 1. This is necessary so that the log odds ratio is finite for all $x \in \mathcal{X}$. Assumption NP.7 is a restriction on the differentiability and smoothness of the propensity score relative to the dimension of **x**. This is a weakening of the smoothness assumption in Hirano, Imbens, and Ridder (2003).

Using the previous assumptions, I obtain the following result:

Proposition 4. Under assumptions NP.2-NP.7, $\|\hat{m} - m_0\|_\infty = o_p(n^{-1/4})$.

The next proof will prove the results for the nonparametric logit sieve estimator as in Hirano, Imbens, and Ridder (2003), but starting from different assumptions. The proof itself is broken up in two parts, first under a set of regularity conditions I prove that the estimator is consistent. Then, I prove that it achieves the desired rate of convergence. First, I prove the following lemma.

Lemma 1. Suppose b,c are arbitrary constants such that $b, c > 0$ and $b \neq c$. Then sign($log(\frac{b}{1+b}) - log(\frac{c}{1+c})$) ≠sign($log(\frac{1}{1+b}) - log(\frac{1}{1+c})$)

Proof. Suppose $b > c$. Then $log(\frac{1}{1+b}) - log(\frac{1}{1+c}) = log(\frac{1+c}{1+b})$. Since $b > c > 0$, then $0 < \frac{1+c}{1+b} < 1$, so $log(\frac{1+c}{1+b}) < 0$. Now, suppose that $log(\frac{b}{1+b}) - log(\frac{c}{1+c}) < 0$. Then $\frac{b}{1+b} < \frac{c}{1+c}$, which implies that $b < c$. This is a contradiction, so $log(\frac{b}{1+b}) - log(\frac{c}{1+c}) > 0$.
Now, suppose $c > b$. Then $log(\frac{1+c}{1+b}) > 0$. If $log(\frac{b}{1+b}) - log(\frac{c}{1+c}) > 0$, then $b > c$. This is a contradiction, so $log(\frac{b}{1+b}) - log(\frac{c}{1+c}) < 0$ □

Proof of Proposition 4:

Proof. Note that

$$\begin{aligned}
|\ell(m, \mathbf{x}_i) - \ell(m', \mathbf{x}_i)| &= \Big|D_i log(\frac{exp(m(\mathbf{x}_i))}{1 + exp(m(\mathbf{x}_i))}) + (1 - D_i) log(\frac{1}{1 + exp(m(\mathbf{x}_i))}) \\
&\quad - D_i log(\frac{exp(m'(\mathbf{x}_i))}{1 + exp(m'(\mathbf{x}_i))}) - (1 - D_i) log(\frac{1}{1 + exp(m'(\mathbf{x}_i))})\Big| \\
&= \Big|D_i[log(\frac{exp(m(\mathbf{x}_i))}{1 + exp(m(\mathbf{x}_i))}) - log(\frac{exp(m'(\mathbf{x}_i))}{1 + exp(m'(\mathbf{x}_i))})] \\
&\quad + (1 - D_i)[log(\frac{1}{1 + exp(m(\mathbf{x}_i))}) - log(\frac{1}{1 + exp(m'(\mathbf{x}_i))})]\Big|
\end{aligned}$$

By the preceding lemma,

$$
\begin{aligned}
&\Big|D_i[log(\frac{exp(m(\mathbf{x}_i))}{1+exp(m(\mathbf{x}_i))}) - log(\frac{exp(m'(\mathbf{x}_i))}{1+exp(m'(\mathbf{x}_i))})] + (1-D_i)[log(\frac{1}{1+exp(m(\mathbf{x}_i))}) - log(\frac{1}{1+exp(m'(\mathbf{x}_i))})]\Big| \\
&\leq \Big|[log(\frac{exp(m(\mathbf{x}_i))}{1+exp(m(\mathbf{x}_i))}) - log(\frac{exp(m'(\mathbf{x}_i))}{1+exp(m'(\mathbf{x}_i))})] - [log(\frac{1}{1+exp(m(\mathbf{x}_i))}) - log(\frac{1}{1+exp(m'(\mathbf{x}_i))})]\Big| \\
&= |m(\mathbf{x}_i) - m'(\mathbf{x}_i)|
\end{aligned}
$$

Then, $sup_{m,m'\in\mathcal{H}:||m-m'||\leq\delta}|\ell(m,\mathbf{x}_i) - \ell(m',\mathbf{x}_i)| \leq \delta$. Hence, Condition (ii) of Theorem 3.5M in Chen (2007) is satisfied. Then by Condition 3.5M, $m_n \xrightarrow{p} m_0$ under $||\cdot||_\infty$.

Now, to prove the second part of the theorem consider the ℓ^2 metric defined by $||m - m_0||_{p,2} = \sqrt{E\,[h(\mathbf{x}_i) - m_0(\mathbf{x}_i)]^2}$. This metric will be used to take advantage of inequalities in relation to $||\cdot||_\infty$ to ultimately find the desired rate of uniform convergence. Suppose $||m - m_0||_{p,2} \leq \epsilon^2$. Note that by the mean value theorem, $\ell(m,\mathbf{x}_i) - \ell(m_0,\mathbf{x}_i) = \frac{d\ell(\tilde{m},\mathbf{x}_i)}{dh}[m - m_0]$, where $\tilde{m}$ lies between m and m_0. Condition 3.7 of Chen (2007) is satisfied by the preceding inequality in the first part of this proof. Lemma 2 in Chen and Shen (1998) implies that $||m - m_0||_\infty \leq C_1||m - m_0||_{p,2}^{2a/(2a+d)}$, where C_1 is a constant and $C_1 > 0$. Then condition 3.8 of Chen (2007) is satisfied with $sup_{||m-m_0||\leq\delta}|\ell(m,\mathbf{x}_i) - \ell(m_0,\mathbf{x}_i)| \leq \delta^{2a/(2a+d)}C_1$. Then by Theorem 3.2 in Chen (2007), $||m - m_0||_{p,2} = O_p(\epsilon_n)$, with $\epsilon = max\{\delta_n, |m_0 - \sigma_n m_0|\}$. Let $u_m = sup_{m\in\mathcal{H}_n}||h||_\infty$, and $||m||_\infty = sup_{\mathbf{x}_i\in X}|m(\mathbf{x}_i)|$. Then for all $0 < \frac{\epsilon}{C_1^2} \leq \delta < 1, log(\frac{\epsilon}{C_1^2}, \mathcal{H}_n, ||\cdot||_\infty) \leq const \cdot k_n \cdot log(1 + \frac{4u_m}{\epsilon})$ by Lemma 2.5 in van der Geer (2000), where $k_n \uparrow$ as $n\to\infty$, but $\frac{k_n}{n}\to 0$. Then,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{H_\square(\epsilon,\mathcal{F}_n,||\cdot||)}d\epsilon \\
&\leq \frac{1}{\sqrt{n}\delta_n^2}\int_{b\delta_n^2}^{\delta_n}\sqrt{k_n\cdot log\left(1+\frac{4u_\pi}{\epsilon}\right)}d\epsilon \\
&\leq \frac{1}{\sqrt{n}\delta_n^2}\sqrt{k_n}\int_{b\delta_n^2}^{\delta_n} log\left(1+\frac{4u_\pi}{\epsilon}\right)d\epsilon \\
&\leq \frac{1}{\sqrt{n}\delta_n^2}\sqrt{k_n}\delta_n \\
&\leq const
\end{aligned}
$$

Then $\delta_n \asymp \sqrt{\frac{k_n}{n}}$, and $||\sigma_k m_0 - m_0||_\infty = O(k_n^{-a/d})$ by Lorentz (1966). Let $\delta_n \asymp ||\sigma_k m_0 - m_0||_\infty$. Then the optimal rate is with $k_n = o(n^{1/(2a+d)})$. Then $||m_n - m_0||_{p,2} = O_p(n^{-a/(2a+d)})$. Now, note that $||m_n - m_0||_{p,2}^{2a/(2a+d)} = o_p(n^{-2a^2/(2a+d)^2})$. Since $||m - m_0||_\infty \leq C_1 ||m - m_0||_{p,2}^{2a/(2a+d)}$ and $||m - m_0||_\infty = o_p(1)$, then $||m - m_0||_\infty = o_p(n^{(-2a^2/(2a+d)^2)}) = o_p(n^{-1/4})$. $\square$

Then I can also show, as in Hirano, Imbens, and Ridder (2003), that

$$sup_{\mathbf{x}\in\mathcal{X}}|\hat{\pi}(\mathbf{x}) - p(\mathbf{x})| = sup_{\mathbf{x}\in\mathcal{X}}|L(\hat{m}(\mathbf{x}) - L(m_0(\mathbf{x})| \lesssim sup_{\mathbf{x}\in\mathcal{X}}|\hat{m}(\mathbf{x}) - m_0(\mathbf{x})| = O(k_n^{-a/d}) = o_p(n^{-1/4})$$

The next proof will tackle the case for nonparametric estimation of the conditional CDF. I need to establish four lemmas before proving Proposition 5.

The next theorem concerns the asymptotic behavior of the estimator of $\hat{P}(\Delta Y_{0t} \leq y|X)$. This estimator is a kernel density estimator, though a sieve estimator could also be chosen. The estimator that I have chosen is based upon the estimator of Li and Racine (2008). The assumptions that are needed include, (from Li and Racine (2008))

Assumption C.1. Both $\mu(x)$ and $F(y|x)$ have continuous second-order partial derivatives with respect to x^c, where x^c denotes the vector of continuous random variables. For fixed values of y and x, $\mu(x) > 0, 0 < F(y|x) < 1$

Assumption C.2. $w(\cdot)$ is a symmetric, bounded, and compactly supported density function, and $w(\cdot)$ is a Lipschitz function on the compact set D.

Assumption C.3. As $n \to \infty, h_s \to \infty$ for $s = 1..., q, \lambda_s \to 0$ for $s = 1, ..., r$, and $(nh_1...h_q) \to \infty$, and as $n \to \infty, h_0 \to 0$.

Assumption C.4. $F(y|x)$ is twice continuously differentiable in (y, x^c).

Let $|h| = \sum_{i=1}^q h_s$, $|\lambda| = \sum_{i=1}^r \lambda_s$, where $0 \leq \lambda_s \leq 1$ and $W_h(X_i^c, x_i^c) = \prod_{s=1}^q h_s^{-1} w((X_{is}^c - x_s^c)/h_s)$. Let $\hat{\mu}(x) = n^{-1} \sum_{i=1}^n W_h(X_i^c, x_i^c)$ Let $\tilde{F}(y|x^c) = \frac{n^{-1}\sum_{i=1}^n G(\frac{y-Y_i}{h_0}) W_h(X_i^c, x_i^c)}{\hat{\mu}(x)}$. $G(\cdot)$ is the distribution function with corresponding density function $w(\cdot)$. h_s is the bandwidth associated with the continuous variable x_s^c, and h_0 is the bandwidth associated with Y_i[7].

We then have the following theorem:

Proposition 5. Suppose Assumptions C.1-C.4 hold and $h_0 = h$. Then, $sup_{x \in D}|\tilde{F}(y|x^c) - F(y|x^c)| = O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h^2)$.

The restriction that $h_0 = h$ will be used to achieve the rate of convergence of $o_p(n^{-1/4})$. I need to establish four lemmas before proving Proposition 5.

The following lemma is similar to Lemma 1 in Li and Racine (2008).

Lemma 2. Under assumptions C.1-C.3, $E[\hat{\mu}(x)] = \mu(x) + O(|h^2|)$

Proof.

$$
\begin{aligned}
E[\hat{\mu}(x)] &= \int \mu(x_i^c) W\left(\frac{X_i^c - x_i^c}{h}\right) dx_i^c (nh_1 \ldots h_q)^{-1} \\
&= \int \mu(x_i^c) k\left(\frac{x_{i1} - x_1}{h_1}\right) \times \cdots \times k\left(\frac{x_{iq} - x_q}{h_q}\right) \\
&= \int \mu(x_i^c + hv) k(v) dv \\
&= \int [\mu(x_i^c) + \sum_{s=1}^{q} \mu_s(x_i^c) h_s v_s + \frac{1}{2} \sum_{s=1}^{q} \sum_{\ell=1}^{q} \mu(x^c) h_s h_\ell v_s v_\ell] k(v) dv + O(|h|^3) \\
&= \mu(x^c) + \frac{\kappa}{2} \sum_{s=1}^{q} \mu_{ss}(x^c) h_s^2 + O(|h|^3) \\
&= \mu(x^c) + O(|h|^2)
\end{aligned}
$$

□

where $\kappa = \int v^2 k(v) dv$.

Lemma 3. Under Assumptions C.1-C.4, $E[\hat{\mu}(x)\tilde{F}(y|x^c)] = \mu(x)F(y|x^c) + \mu(x) \sum_{i=1}^{q} h_s^2 B_s(y, x) + o(|h|^2) + o(h_0^2)$

[7]In principle, the estimator here could be the estimator of Li and Racine (2008), where the covariates can be discrete and ordered; however, in order to cite particular theorems from Rothe and Firpo (2013) I am only considering an estimator that allows for continuous covariates, though as noted by Rothe and Firpo (2019), the results could be modified to allow for discrete covariates.

Proof. See Theorem 6.2 (i) in Li and Racine (2007) □

Now, the rate of uniform convergence proof largely follows Masry (1996). Furthermore, since by Li and Racine (2008) it is shown that at the optimal (to minimize the integrated mean square error) occurs when $h_1, \ldots, h_q$ all converge to 0 at the same rate. I will denote this common h by h_{min}.

Lemma 4. Under assumptions C.1-C.4, $sup_{x \in D} |\hat{\mu}(x) - \mu(x)| = O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(|h|^2)$.

Proof. Note that by the Triangle Inequality,

$$\begin{aligned}|\hat{\mu}(x) - \mu(x)| &= |\hat{\mu}(x) - \mu(x) - E[\hat{\mu}(x)] + E[\hat{\mu}(x)]| \\ &\leq |\mu(x) - E[\hat{\mu}(x)]| + |E[\hat{\mu}(x)] - \hat{\mu}(x)|\end{aligned}$$

By Lemma 2 and Lemma 3, $|\mu(x) - E[\hat{\mu}(x)]| = O(|h|^2)$. Then it is sufficient to find the rate for $|E[\hat{\mu}(x)] - \hat{\mu}(x)|$. Since D is compact, it can be covered by a finite number $L = L(n)$ of cubes $I_k = I_{n,k}$ with centers $x_k = x_{n,k}$ having sides of length ℓ_n for $k = 1, ..., L(n)$. Clearly $\ell_n = constant/L^{1/d}(n)$. Since D is compact, write

$$\begin{aligned}sup_{x \in D} |E[\hat{\mu}(x)] - \hat{\mu}(x)| &= max_{1 \leq k \leq L_n} sup_{x \in D \cap I_k} |E[\hat{\mu}(x)] - \hat{\mu}(x)| \\ &\leq max_{1 \leq k \leq L_n} sup_{x \in D \cap I_k} |\hat{\mu}(x) - \hat{\mu}(x_{k,n})| \\ &+ max_{1 \leq k \leq L_n} |E[\hat{\mu}(x_{k,n})] - \hat{\mu}(x_{k,n})| \\ &+ max_{1 \leq k \leq L_n} sup_{x \in D \cap I_k} |E[\hat{\mu}(x_{k,n})] - E[\hat{\mu}(x)]| \\ &:= Q_1 + Q_2 + Q_3\end{aligned}$$

Since each kernel is Lispschitz, and the product of Lipschitz functions is a Lipschitz function,

$$\begin{aligned}Q_1 &= |\hat{\mu}(x) - \hat{\mu}(x_{k,n})| \\ &\leq |W_h(X_i^c, x^c) - W_h(X_i^c, x_{k,n}^c)|\end{aligned}$$

$$\leq (C_2/h_{min}^{q+1}) sup_{x \in D \cap I_k} |x - x_{k,n}|$$
$$\leq C_2 \ell_n / h_{min}^{q+1}.$$

Let $\ell_n = (ln(n))^{1/2} h^{(q+2)/2} / n^{1/2}$. Then $Q_1 = O((ln(n)/(nh^q))^{1/2})$. Similarly, $Q_3 = O((ln(n)/(nh^q))^{1/2})$ Now let $W_n(x) = \hat{\mu}(x) - E[\hat{\mu}(x)] = \sum_{i=1} Z_{n,i}$ where,

$$Z_{n,i} = (nh_{min}^q)^{-1} [W_h(X_i^c, x_i^c)] - E[W_h(X_i^c, x_i^c)]]$$

For $\eta > 0$, we have

$$\begin{aligned}
P[Q_2 > \eta] &\leq P[max_{1 \leq k \leq L_n} W_n(x_{k,n}) > \eta] \\
&\leq P[W_n(x_{1,n}) > \eta \text{ or } W_n(x_{2,n}) > \eta, \ldots, \text{ or } W_n(x_{L(n),n}) > \eta] \\
&\leq P[W_n(x_{1,n}) > \eta + W_n(x_{2,n}) > \eta, \ldots, +W_n(x_{L(n),n}) > \eta] \\
&\leq sup_{x \in S} P[|W_n(x)| > \eta]
\end{aligned}$$

Since $\hat{\mu}(\cdot)$ is bounded, and letting $A = sup_{x \in D} |\hat{\mu}(x)|$, we have $|Z_{n,i}| \leq 2A/nh_{min}^q$ for all $i = 1, ..., n$. Define $\gamma_n = (nh_{min}^q ln(n))^{1/2}$. Then $\gamma_n |Z_{n,i}| \leq 2A(ln(n))/(nh_{min}^q]^{1/2} \leq 1/2$ for all $i = 1, ..., n$ for n sufficiently large. Using the inequality $e^x \leq 1+x+x^2$ for $|x| \leq 1/2$, we have $e^{\gamma_n Z_{n,i}} \leq 1+\gamma_n Z_{n,i}+\gamma_n^2 Z_{n,i}^2$. Hence, $E[e^{\gamma_n Z_{n,i}}] \leq 1 + \gamma_n^2 E[Z_{n,i}^2] \leq e^{E[\gamma_n^2 Z_{n,i}^2]}$. Then,

$$\begin{aligned}
P[|W_n(x)| > \eta] &= P[|\sum_{i=1}^{n} Z_{n,i}| > \eta] \\
&= P[\sum_{i=1}^{n} Z_{n,i} > \eta] + P[\sum_{i=1}^{n} Z_{n,i} < -\eta] \\
&\leq P[\sum_{i=1}^{n} Z_{n,i} > \eta] + P[-\sum_{i=1}^{n} Z_{n,i} > \eta] \\
&\leq E[e^{\gamma_n \sum_{i=1}^{n} Z_{n,i}}] + E[e^{-\gamma_n \sum_{i=1}^{n} Z_{n,i}}] \\
&\leq 2e^{-\gamma_n} e^{\gamma_n^2 \sum_{i=1}^{n} E(Z_{n,i}^2)}
\end{aligned}$$

$$\leq 2e^{-\gamma_n} e^{A\gamma_n^2/(nh^q_{min})}$$

Then $sup_{x\in D}P[|W_n(x)| > \eta] \leq 2e^{-\gamma_n\eta+\frac{A\gamma_n^2}{nh^q_{min}}}$. Let $\gamma_n\eta = C_3 ln(n)$. Choose $\gamma_n = [(nh^q_{min}ln(n)]^{1/2}$. Then $-\gamma_n\eta/\alpha + A\gamma_n^2/(nh^q_{min}) = -C_3 ln(n) + Aln(n) = -\alpha ln(n)$, where $\alpha = C_3 - A$. Since $sup_{x\in D}P[|W_n(x)| > \eta] \leq 2e^{-\gamma_n\eta+\frac{A\gamma_n^2}{nh^q_{min}}}$ and $P[Q_2 > \eta] \leq L(n)sup_{x\in D}P[|W_n(x)| > \eta]$, then $P[Q_2 > \eta_n] \leq 2L(n)/n^{\alpha}$. Choose C_3 sufficiently large and $L(n)$ such that $\sum_{n=1}^{\infty} P[|Q_2/\eta_n| > 1] \leq 4\sum_{n=1}^{\infty} L(n)/n^{\alpha} < \infty$. Then by the Borel-Cantelli lemma, $Q_2 = O_p((ln(n))^{1/2}/(nh^q_{min})^{1/2})$. □

Similarly, by Lemma 3, and by a similar result to Lemma 4, $sup_{x\in D}|\hat{\mu}(x)\tilde{F}(y|x^c) - \mu(x)F(y|x^c)| = O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)$. Then I have the following theorem,

Proof of Proposition 5

Proof. Note that $\tilde{F}(y|x^c) = \frac{\hat{\mu}(x)\tilde{F}(y|x^c)}{\hat{\mu}(x)} = \frac{\hat{\mu}(x)\tilde{F}(y|x^c)/\mu(x)}{\hat{\mu}(x)/\mu(x)}$. By Lemma 4,

$$sup_{x\in D}|\hat{\mu}(x) - \mu(x)| = O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(|h|^2)$$

Then, by Lemma 4

$$\begin{aligned} sup_{x\in D}|\frac{\hat{\mu}(x)}{\mu(x)} - 1| &= sup_{x\in D}|\frac{\hat{\mu}(x) - \mu(x)}{\mu(x)}| \\ &\leq \frac{O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(|h|^2)}{inf_{x\in D}\mu(x)} \\ &= O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(|h|^2) \end{aligned}$$

Similarly, since

$$sup_{x\in D}|\hat{\mu}(x)\tilde{F}(y|x^c) - \mu(x)F(y|x^c)| = O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)$$

Then,

$$sup_{x\in D}|\frac{|\hat{\mu}(x)\tilde{F}(y|x^c)}{\mu(x)} - F(y|x^c)| \leq \frac{O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)}{inf_{x\in D}\mu(x)}$$
$$\leq O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)$$

Then,

$$\begin{aligned}
\tilde{F}(y|x^c) &= \frac{\hat{\mu}(x)\tilde{F}(y|x^c)/\mu(x)}{\hat{\mu}(x)/\mu(x)} \\
&= \frac{F(y|x^c) + O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)}{1 + O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)} \\
&= \frac{F(y|x^c) + O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h_0^2) + O_p(|h|^2)}{1 + O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(|h|^2)} \\
&= F(y|x^c) + O_p(\frac{ln(n)^{1/2}}{(nh^q)^{1/2}}) + O_p(h^2)
\end{aligned}$$

□

B Proofs of Theorems and Propositions in Main Text

The first proof is of the identification result in Theorem 1.

Proof of Theorem 1

Proof. Note that by Theorem 1 in Callaway and Li (2019), the first portion of the result is proven. All that remains is to show that $F_{\Delta Y_{0t}|D=1}(\delta) = E\left[\left(\frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)}\right)\mathbb{1}\{\Delta Y_t \leq \delta\} - \left(\frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)} - \frac{D}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|X)\right]$ if $\pi(X) = p(X)$ a.c., or $\tilde{P}(\Delta Y_{0t} \leq \delta|X) = P(\Delta Y_{0t} \leq \delta|X)$ a.c. Suppose $\pi(X) = p(X)$ a.c. Then,

$$E\left[\left(\frac{(1-D)p(X)}{p(1-p(\mathbf{x}))}\right)\mathbb{1}_{\Delta Y_t \leq \delta} - \left(\frac{(1-D)p(X)}{p(1-p(X))}) - \frac{D}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})\right]$$
$$= E\left[\frac{p(x)E[(1-D)\mathbb{1}_{\Delta Y_t \leq \delta|D=0,X}}{p}\right] - E\left[\frac{p(X)\tilde{P}(\Delta Y_{0t}|\mathbf{x}, D=0)}{p}\right] + E\left[\frac{p(X)\tilde{P}(\Delta Y_{0t} \leq \delta|X, D=1)}{p}\right]$$

$$
\begin{aligned}
&= E\left[\frac{p(X)P(\Delta Y_{0t} \le \delta|X, D=0)}{p}\right] - E\left[\frac{p(X)\tilde{P}(\Delta Y_{0t} \le \delta|X, D=1)}{p}\right] + E\left[\frac{p(X)\tilde{P}(\Delta Y_{0t} \le \delta|X, D=1)}{p}\right] \\
&= E\left[\frac{p(X)P(\Delta Y_{0t} \le \delta|X, D=1)}{p}\right] \\
&= E\left[\frac{P(\Delta Y_{0t} \le \delta, D=1|X)}{p}\right] \\
&= P(\Delta Y_{0t} \le \delta|D=1) \\
&= F_{\Delta Y_{0t}|D=1}(\delta)
\end{aligned}
$$

Now, suppose $\pi(X) \neq p(X)$ a.c. and $\tilde{P}(\Delta Y_{0t} \le \delta|X) = P(\Delta Y_{0t} \le \delta|X)$ a.c. Then,

$$
\begin{aligned}
&E\left[\left(\frac{(1-D)\pi(X)}{p(1-\pi(X))}\right)\mathbb{1}_{\Delta Y_t \le \delta} - \left(\frac{(1-D)\pi(X)}{p(1-\pi(X))}) - \frac{D}{p}\right)P(\Delta Y_{0t} \le \delta|X)\right] \\
&= E\left[\frac{p(X)P(\Delta Y_{0t} \le \delta|X, D=1)}{p}\right] - E\left[\frac{(1-p(X))\pi(X)P(\Delta Y_{0t} \le \delta|X, D=0)}{p(1-\pi(X))}\right] + E\left[\frac{(1-p(X))\pi(X)P(\Delta Y_{0t} \le \delta|X, D=0)}{p(1-\pi(X))}\right] \\
&= E\left[\frac{P(\Delta Y_{0t} \le \delta, D=1|X)}{p}\right] \\
&= P(\Delta Y_{0t} \le \delta|D=1) \\
&= F_{\Delta Y_{0t}|D=1}(\delta)
\end{aligned}
$$

□

The next proof holds in either the parametric or nonparametric subcase, though this proof is for a parametric submodel, while the nonparametric submodel will proceed similarly. For the purpose of estimation of $F_{\Delta Y_{0t}|D=1}(\delta)$, only the period of treatment and the period prior needs to be considered. If there are additional pre-treatment and post-treatment periods, they are not relevant to the density of the data that is used to estimate $F_{\Delta Y_{0t}|D=1}(\delta)$. The proof itself is similar to a result in Sant'Anna and Zhao (2020).

Proof of Theorem 2:

Proof. The density of $(y_t(1), y_t(0), y_{t-1}(0), d, \mathbf{x})$ with respect to some sigma-finite measure on $\mathcal{L} \in$

$\mathbb{R}^3 \times \{0,1\} \times \mathbb{R}^k$ is given by

$$\bar{f}(y_t(1), y_t(0), y_{t-1}(0), d, \mathbf{x}) = \bar{f}(y_t(1), y_t(0), y_{t-1}(0)|D = 1, \mathbf{x})^d p(\mathbf{x})^d \bar{f}(y_t(1), y_t(0), y_{t-1}(0)|D = 0, x)^{1-d}(1 - p(x))^{1-d} f(\mathbf{x})$$

The density of the observed data is,

$$f(y_t, y_{t-1}, d, x) = f_1(y_t, y_{t-1}|D = 1, \mathbf{x})^d p(\mathbf{x})^d f_0(y_t, y_{t-1}|D = 0, \mathbf{x})^{1-d}(1 - p(\mathbf{x}))^{1-d} f(\mathbf{x})$$

where

$$\begin{aligned} f_1(\cdot, \cdot|D = 1, \mathbf{x}) &= \int \bar{f}(\cdot, y_t(0), \cdot|D = 1, \mathbf{x}) dy_t(0) \\ f_0(\cdot, \cdot|D = 0, \mathbf{x}) &= \int \bar{f}(y_t(1), \cdot, \cdot|D = 0, \mathbf{x}) dy_t(1) \end{aligned}$$

Consider a parametric submodel indexed by a parameter θ,

$$f_\theta(y_t, y_{t-1}, d, \mathbf{x}) = f_{1,\theta}(y_t, y_{t-1}|D = 1, \mathbf{x})^d p_\theta(\mathbf{x})^d f_{0,\theta}(y_t, y_{t-1}|D = 0, \mathbf{x})^{1-d}(1 - p_\theta(\mathbf{x}))^{1-d} f_\theta(\mathbf{x})$$

which equals $f(y_t, y_{t-1}, d, \mathbf{x})$ when $\theta = \theta_0$. The score is

$$s_\theta(y_t, y_{t-1}, d, \mathbf{x}) = d s_{1\theta}(y_t, y_{t-1}|D = 1, \mathbf{x}) + (1 - d) s_{0\theta}(y_t, y_{t-1}|D = 0, \mathbf{x}) + \frac{d - p_\theta(\mathbf{x})}{p_\theta(\mathbf{x})(1 - p_\theta(\mathbf{x}))} \dot{p}_\theta(\mathbf{x}) + t_\theta(\mathbf{x})$$

where, for $d = 0, 1$

$$s_{d\theta}(y_t, y_{t-1}|D = d, \mathbf{x}) = \frac{d}{d\theta} log f_{d,\theta}(y_t, y_{t-1}|D = d, \mathbf{x}), \;\; \dot{p}_\theta(\mathbf{x}) \frac{d}{d\mathbf{x}} p_\theta(\mathbf{x}), \text{ and } t_\theta(\mathbf{x}) = \frac{d}{d\theta} log f_\theta(\mathbf{x})$$

Then the tangent space is

$$\mathcal{F} = \{d s_1(y_t, y_{t-1}|D = 1, \mathbf{x}) + (1 - d) s_0(y_t, y_{t-1}|D = 0, \mathbf{x}) + a(x)(d - p(\mathbf{x})) + t(\mathbf{x})\}$$

where $\iint s_d(y_t, y_{t-1}|D = d, \mathbf{x}) f_d(y_t, y_{t-1}|D = d, \mathbf{x}) dy_y dy_{t-1} = 0 \; \forall x, d = 0, 1$, $\int t(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = 0$ and $a(\mathbf{x})$ is any square integrable function of $\mathbf{x}$. Under the assumption that $\Delta Y_{0t} \perp\!\!\!\perp D|X$, note that $\tau = F_{\Delta Y_t|D=1}(\delta)$,

$$\tau = E[E[\mathbb{1}_{\Delta Y_{0t} \leq \delta}|D = 1, X]|D = 1] = E[E[\mathbb{1}_{\Delta Y_{0t} \leq \delta}|D = 0, X]|D = 1]$$

For the parametric submodel under consideration, I note that

$$\tau(\theta) = \frac{\iiint \mathbb{1}_{y_t \leq \delta + y_{t-1}} p_\theta(\mathbf{x}) f_{0,\theta}(y_t, y_{t-1}|D = 0, \mathbf{x}) f_\theta(\mathbf{x}) dy_3 dy_2 d\mathbf{x}}{\int p_\theta(\mathbf{x}) f_\theta(\mathbf{x}) d\mathbf{x}}$$

Then,

$$\begin{aligned}
\frac{\partial \tau(\theta_0)}{\partial \theta} &= \frac{\iiint \mathbb{1}_{y_t \leq \delta + y_{t-1}} p(\mathbf{x}) s_0(y_t, y_{t-1}|D = 0, \mathbf{x}) f_0(y_t, y_{t-1}|D = 0, \mathbf{x}) f(\mathbf{x}) dy_3 dy_2 d\mathbf{x}}{p} \\
&+ \frac{\int P(\Delta Y_{0t} \leq \delta|\mathbf{x}, D = 0) p(\mathbf{x}) t(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}}{p} + \frac{\int P(\Delta Y_{0t} \leq \delta|\mathbf{x}, D = 0) p(\mathbf{x}) \dot{p}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}}{p} \\
&- \frac{\tau[\int \dot{p}(\mathbf{x}) + p(\mathbf{x}) t(\mathbf{x})] f(\mathbf{x}) d\mathbf{x}]}{p}
\end{aligned}$$

Let the initial choice of an influence function be,

$$\begin{aligned}
F_\tau(Y_t, Y_{t-1}, D, X) &= \frac{(1-D)p(X))}{p(1-p(X))} \mathbb{1}_{\Delta Y_t \leq \delta} - \frac{(1-D)p(X)}{p(1-p(X))} P(\Delta Y_t \leq \delta|X, D = 0) \\
&+ \frac{D - p(X)}{p} P(\Delta Y_t \leq \delta|X, D = 0) + \frac{p(X)}{p} P(\Delta Y_t \leq \delta|X, D = 0) - \frac{D}{p}\tau \\
&= \frac{(1-D)p(X)}{p(1-p(X))} \mathbb{1}_{\Delta Y_t \leq \delta} - \frac{(1-D)p(X)}{p(1-p(X))} P(\Delta Y_{0t} \leq \delta|X) \\
&+ \frac{D}{p} P(\Delta Y_{0t} \leq \delta|X) - \frac{D}{p}\tau
\end{aligned}$$

Note that for the parametric submodel with score $s_\theta(y_1, y_0, d, \mathbf{x})$, I can conclude that τ is a differentiable parameter since

$$\frac{\partial \tau(\theta_0)}{\partial \theta} = E[F_\tau(Y_t, Y_{t-1}, D, X) s_\theta(Y_1, Y_0, D, X)]$$

Since $F_\tau \in \mathcal{F}$, then by Theorem 3.1 of Newey (1990), $F_\tau(Y_t, Y_{t-1}, D, X)$ is the efficient influence function for $F_{\Delta Y_{0t}|D=1}(\delta)$. $\square$

Proof of Theorem 3:

Proof. **Consistency of estimator, nonparametric case**:

$$\hat{F}_{\Delta Y_{0t}|D=1}(\delta) = n^{-1}\sum_{i=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_t \leq \delta\} - \left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})\right]$$

Suppose that $\hat{\pi}(\mathbf{x}) \xrightarrow{p} \pi(\mathbf{x})$ Furthermore, $\sum_{k=1}^{n}\frac{D_k}{n} \xrightarrow{p} p$. Then by the WLLN and the Continuous Mapping Theorem,

$$n^{-1}\sum_{i=1}^{n}\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_{ti} \leq \delta\} \xrightarrow{p} E\left(\frac{1-D}{p}\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\mathbb{1}\{\Delta Y_t \leq \delta\}\right)$$

Then,

$$n^{-1}\sum_{i=1}^{n}\left[-\left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})\right] \xrightarrow{p} E\left[-\left(\frac{1-D_i}{p}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} - \frac{D_i}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})\right]$$

This implies that $\hat{F}_{\Delta Y_{0t}|D=1}(\delta) \xrightarrow{p} E\left[\frac{1-D}{p}\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\mathbb{1}\{\Delta Y_t \leq \delta\}\right] + E\left[-\left(\frac{1-D_i}{p}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} - \frac{D_i}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|\mathbf{x})\right]$. If $\pi(X) = p(X)$ a.c. or $\tilde{P}(\Delta Y_{0t} \leq \delta|X) = P(\Delta Y_{0t} \leq \delta|X)$ a.c., then by the previous theorem

$$E\left[\frac{1-D}{p}\frac{\pi(X)}{1-\pi(X)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right] + E\left[-\left(\frac{1-D_i}{p}\frac{\pi(X)}{1-\pi(X)} - \frac{D_i}{p}\right)P(\Delta Y_{0t} \leq \delta|X)\right] = F_{\Delta Y_{0t}|D=1}(\delta)$$

$$\hat{F}_{\Delta Y_{0t}|D=1}(y) - F_{\Delta Y_{0t}|D=1}(y) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\right)\left[\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i)\right]\right] - F_{\Delta Y_{0t}|D=1}(y)$$

The next proof follows partly from the proof of Theorem 2(b) in Rothe and Firpo (2019). In particular, the object is to expand the doubly robust moment condition and demonstrate that each term converges in probability to 0 at the desired rate. In the parametric case, the proof is very similar to Sant'Anna and Zhao (2020). The proof in nonparametric case is also similar to Fan et al. (2016) when the nuisance function is estimated using a sieve approach. Now, I will expand

$$\hat{F}_{\Delta Y_{0t}|D=1}(y) - F_{\Delta Y_{0t}|D=1}(y) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\hat{\pi}(\mathbf{x}_i)}{1-\hat{\pi}(\mathbf{x}_i)} - \frac{D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\right)\left[\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i)\right]\right] - F_{\Delta Y_{0t}|D=1}(y)$$

Let

$$\begin{aligned}
\psi_i &= \left(\frac{1-D_i}{p}\frac{\hat{\pi}(\mathbf{x}_i)}{(1-\hat{\pi}(\mathbf{x}_i))}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D_i}{p}\frac{1}{(1-\hat{\pi}(\mathbf{x}_i))^2} - \frac{D_i}{\hat{p}}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)] - \frac{D_i}{\hat{p}}F_{\Delta Y_{0t}|D=1}(y) \\
\psi_i^2 &= \left(-\frac{1-D_i}{p}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} + \frac{D_i}{p}\right) \\
\psi_i^{22} &= 0 \\
\psi_i^1 &= \left(\frac{1-D_i}{p}\frac{1}{(1-\hat{\pi}(\mathbf{x}_i))^2}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D_i}{p}\frac{1}{(1-\hat{\pi}(\mathbf{x}_i))^2}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)] \\
\psi_i^{11} &= \left(\frac{1-D_i}{p}\frac{2\hat{\pi}(\mathbf{x}_i)}{(1-\hat{\pi}(\mathbf{x}_i))^3}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{1-D_i}{p}\frac{2\hat{\pi}(\mathbf{x}_i)}{(1-\hat{\pi}(\mathbf{x}_i))^3}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)] \\
\psi_i^{12} &= \left(\frac{1-D_i}{p}\frac{1}{(1-\hat{\pi}(\mathbf{x}_i))^2}\right) \\
\psi_i^{13} &= \left(\frac{D_i-1}{\hat{p}^2}\frac{1}{(1-\pi(\mathbf{x}_i))^2}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{D_i-1}{\hat{p}^2}\frac{1}{(1-\pi(\mathbf{x}_i))^2}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)] \\
\psi_i^{23} &= \left(-\frac{D_i-1}{\hat{p}^2}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} - \frac{D_i}{\hat{p}^2}\right) \\
\psi_i^3 &= \left[\left(\frac{D_i-1}{\hat{p}^2}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{D_i-1}{\hat{p}^2}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} + \frac{D_i}{\hat{p}^2}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)]\right] + \frac{D_i}{\hat{p}^2}F_{\Delta Y_{0t}|D=1}(y) \\
\psi_i^{33} &= \left[\left(\frac{2(1-D_i)}{\hat{p}^3}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)}\right)\mathbb{1}\{\Delta Y_t \leq y\} - \left(\frac{2(1-D_i)}{\hat{p}^3}\frac{\pi(\mathbf{x}_i)}{1-\pi(\mathbf{x}_i)} - \frac{2D_i}{\hat{p}^3}\right)[P(\Delta Y_{0t} \leq y|\mathbf{x}_i)]\right] - 2\frac{D_i}{\hat{p}^3}F_{\Delta Y_{0t}|D=1}(y) \\
\Psi_n(\hat{p}, \hat{\pi}, \hat{P}) &= \frac{1}{n}\sum_{i=1}^{n}\psi(D_i, \hat{\pi}(x_i), \hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i), \hat{p})
\end{aligned}$$

Then,

$$\Psi_n(\hat{p}, \hat{\pi}, \hat{P}) - \Psi_n(p, \pi, P) = \frac{1}{n}\sum_{i=1}^{n}\psi_i^1(\hat{\pi}(\mathbf{x}_i) - \pi(\mathbf{x}_i)) + \frac{1}{n}\sum_{i=1}^{n}\psi_i^2(\hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}_i) - P(\Delta Y_{0t} \leq y|\mathbf{x}_i))$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{3}(\hat{p}-p)+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{11}(\hat{\pi}(\mathbf{x}_i)-\pi(\mathbf{x}_i)^2\\
&+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{12}(\hat{P}(\Delta Y_{0t}\leq y|\mathbf{x}_i)-P(\Delta Y_{0t}\leq y|\mathbf{x}_i))(\hat{\pi}(\mathbf{x}_i)-\pi(\mathbf{x}_i))\\
&+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{13}(\hat{\pi}(\mathbf{x}_i)-\pi(\mathbf{x}_i))(\hat{p}-p)\\
&+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{23}(\hat{P}(\Delta Y_{0t}\leq y|\mathbf{x}_i)-P(\Delta Y_{0t}\leq y|\mathbf{x}_i))(\hat{p}-p)\\
&+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{22}(\hat{P}(\Delta Y_{0t}\leq y|\mathbf{x}_i)-P(\Delta Y_{0t}\leq y|\mathbf{x}_i))^2+\frac{1}{n}\sum_{i=1}^{n}\psi_i^{33}(\hat{p}-p)^2\\
&+O_p(||\hat{\pi}-\pi||_\infty^3)+O_p(||\hat{P}-P||_\infty^3)+O_p(||\hat{p}-p||_\infty^3)
\end{aligned}
$$

It remains to show that each term is $o_p(n^{1/2})$. Each term outside of the first term converges converges either due to Rothe and Firpo (2019) or due to the fact that $|\hat{\pi}(\mathbf{x}_i)-\pi(\mathbf{x}_i)| = o_p(n^{1/4})$. For the first term, let $\mathbb{G}_n(f_0) = n^{1/2}(\mathbb{P}_n-\mathbb{P})f_0(D,\Delta Y_t,\mathbf{x})$, where $\mathcal{P}_n$ is the empirical measure, $\mathcal{P}$ is the expectation, and

$$
f_0(D,\Delta Y_t,\mathbf{x}) = \frac{(1-D)(\mathbb{1}_{\Delta Y-t\leq y}-P(\Delta Y_{0t}\leq y|\mathbf{x}))}{p(1-\pi(\mathbf{x}))^2}(\hat{\pi}(\mathbf{x}_i)-\pi(\mathbf{x}_i))
$$

Since $sup_{x\in\mathcal{X}}|\hat{\pi}(\mathbf{x})-\pi(\mathbf{x})| \lesssim O(k_n^{-a/d}) = o_p(1)$ by Theorem 3.2 in Chen (2007) and the proof of Proposition 4, then define $\mathcal{F} = \{f_0 : ||\hat{\pi}(\mathbf{x})-\pi(\mathbf{x})||_\infty \leq \delta_n\}$, where $\delta_n = C(k_n^{-a/d})$ for some $C>0$. By Lemma 3 in Rothe and Firpo (2013), $\mathbb{P}f_0(D.\Delta Y_t,\mathbf{x}) = 0$. By the Markov inequality and Corollary 19.35 of Vaart (2000)

$$
\frac{\sum_{i=1}^{n}\psi_i^{1}(\hat{\pi}(x)-\pi(x))}{n^{1/2}} \leq sup_{f_0\in\mathcal{F}}\mathbb{G}_n(f_0) \lesssim J_{[]}(||F_0||_{p,2},\mathcal{F},L_2(p))
$$

where $J_{[]}(||F_0||p,2,\mathcal{F},L_2(p))$ is the bracketing integral, and F_0 is the envelope function. Let $\mathcal{F}_0 = \{f_0 : ||\hat{\pi}(\mathbf{x})-\pi(\mathbf{x})||_\infty \leq C\}$ Since p and $\pi(x)$ is bounded away from 0, then

$$
|f_0(D,\Delta Y_t,\mathbf{x})| \lesssim \delta_n|\mathbb{1}_{\Delta Y-t\leq y}-P(\Delta Y_{0t}\leq y|\mathbf{x})| := F_0
$$

Then since $\mathbb{1}_{\Delta Y-t\leq y} - P(\Delta Y_{0t} \leq y|x))$ is bounded by 1, $||F_0||_{p,2} \leq \delta_n$. Then,

$$logN_{[]}(\epsilon, \mathcal{F}, L_2(p)) \lesssim logN_{[]}(\epsilon, \mathcal{F}_0\delta_n, L_2(p) = logN_{[]}(\epsilon/\delta_n, \mathcal{F}_0, L_2(p) \lesssim logN_{[]}(\epsilon/\delta_n, \Lambda_c^p(\mathcal{X}), L_2(p) \lesssim (\delta_n/\epsilon)^{d/p}$$

where the last inequality follows by Corollary 2.7.2 in Vaart and Wellner (1996). Then,

$$J_{[]}(||F_0||_{p,2}, \mathcal{F}, L_2(p)) \lesssim \int_0^{\delta_n} \sqrt{logN_{[]}(\epsilon, \mathcal{F}, L_2(p))} d\epsilon \lesssim \int_0^{\delta_n} (\delta_n/\epsilon)^{d/a} d\epsilon \xrightarrow{\delta_n \xrightarrow{n\to\infty} 0} 0$$

where the integral converges to zero since $d/a > 2$ by Assumption NP.7. Then $\frac{\sum_{i=1}^n \psi_i^1(\hat{\pi}(x)-\pi(x))}{n^{1/2}} = o_p(1)$.

Consistency of estimator, parametric case:

$$\hat{F}_{\Delta Y_{0t}|D=1}(\delta) = n^{-1}\sum_{i=1}^n \left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^n D_k}{n}} \frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x}_i;\hat{\gamma})}\right)\mathbb{1}\{\Delta Y_t \leq \delta\} - \left(\frac{1-D_i}{\frac{\sum_{k=1}^N D_k}{n}} \frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x}_i;\gamma)} - \frac{D_i}{\frac{\sum_{k=1}^n D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq \delta|X;\hat{\beta})\right]$$

Suppose that $\hat{\gamma} \xrightarrow{p} \gamma^*$ Furthermore, $\sum_{k=1}^n \frac{D_k}{n} \xrightarrow{p} p$. Then by the WLLN and the Continuous Mapping Theorem,

$$n^{-1}\sum_{i=1}^n \left(\frac{1-D_i}{\frac{\sum_{k=1}^n D_k}{n}} \frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x}_i;\hat{\gamma})}\right)\mathbb{1}\{\Delta Y_{0t} \leq \delta\} \xrightarrow{p} E\left(\frac{1-D}{p}\frac{\pi(\mathbf{x}_i;\gamma^*)}{1-\pi(\mathbf{x}_i;\gamma^*)}\mathbb{1}\{\Delta Y_{0t} \leq \delta\}\right)$$

Assume that $\hat{\beta} \xrightarrow{p} \beta^*$. Then,

$$n^{-1}\sum_{i=1}^n \left[-\left(\frac{1-D_i}{\frac{\sum_{k=1}^N D_k}{n}} \frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x}_i;\hat{\gamma})} - \frac{D_i}{\frac{\sum_{k=1}^n D_k}{n}}\right)\hat{P}(\Delta Y_{0t} \leq \delta|X_i;\hat{\beta})\right] \xrightarrow{p} E\left[-\left(\frac{1-D_i}{p}\frac{\pi(\mathbf{x}_i;\gamma^*)}{1-\pi(\mathbf{x}_i;\gamma^*)} - \frac{D_i}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|\mathbf{x}_i)\right]$$

This implies that $\hat{F}_{\Delta Y_{0t}|D=1}(\delta) \xrightarrow{p} E\left[\frac{1-D}{p}\frac{\pi(\mathbf{x};\gamma^*)}{1-\pi(\mathbf{x};\gamma)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right] + E\left[-\left(\frac{1-D_i}{p}\frac{\pi(\mathbf{x}_i;\gamma^*)}{1-\pi(\mathbf{x}_i;\gamma^*)} - \frac{D_i}{p}\right)P(\Delta Y_{0t} \leq \delta)\right]$.
If $\pi(X;\gamma) = p(X;\gamma^*)$ a.c. or $\tilde{P}(\Delta Y_{0t} \leq \delta|X;\beta) = P(\Delta Y_{0t} \leq \delta|X;\beta^*)$ a.c., then by the previous

theorem

$$E\left[\frac{1-D}{p}\frac{\pi(X;\gamma^*)}{1-\pi(X;\gamma^*)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right] + E\left[-\left(\frac{1-D_i}{p}\frac{\pi(X;\gamma^*)}{1-\pi(X;\gamma^*)} - \frac{D_i}{p}\right)P(\Delta Y_{0t} \leq \delta|X)\right] = F_{\Delta Y_{0t}|D=1}(\delta)$$

Note that,

$$\begin{aligned}
&\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(\delta) \\
&\qquad = n^{-1}\sum_{i=1}^{n}\left(\frac{1-D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x}_i;\hat{\gamma})}\right)\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E\left(\frac{1-D}{p}\frac{\pi(\mathbf{x};\gamma)}{1-\pi(\mathbf{x};\gamma)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right) \\
&\qquad - n^{-1}\sum_{i,j=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\hat{\pi}(\mathbf{x}_i;\hat{\gamma})} - \frac{D_i}{\frac{\sum_{k=1}^{n} D_k}{n}}\right)\left[\frac{(1-D_j)(\mathbb{1}\{\Delta\mu_{0t}(\mathbf{x}_i;\hat{\beta})+\Delta\hat{u}_{0tj}\})}{n_{1-D}}\right]\right] - E\left[\left(\frac{1-D_i}{p}\frac{\tilde{p}(\mathbf{x}_i;\gamma)}{1-\pi(\mathbf{x}_i;\gamma)} - \frac{D_i}{p}\right)\tilde{P}(\Delta Y_{0t} \leq \delta|\mathbf{x};\beta)\right] \\
&\qquad = (C\hat{D}F^1 - CDF^1) - (C\hat{D}F^2 - CDF^2)
\end{aligned}$$

Let $w_0(D,\mathbf{x};\hat{\gamma}) = \left(\frac{1-D}{\frac{\sum_{k=1}^{n} D_k}{n}}\frac{\pi(\mathbf{x};\hat{\gamma})}{1-\pi(\mathbf{x};\hat{\gamma})}\right)$.
Then,

$$\begin{aligned}
&\sqrt{n}(C\hat{D}F^1 - CDF^1) \\
&= n^{-1/2}\sum_{i=1}^{n}(w_0(D_i,\mathbf{x}_i;\hat{\gamma})\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&= n^{-1/2}\sum_{i=1}^{n}(\tilde{w}_0(D_i,\mathbf{x}_i;\hat{\gamma})\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&- \sqrt{n}\sum_{i=1}^{n}\left((1-D_i)\frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\pi(\mathbf{x};\hat{\gamma})}\right) - E\left[(1-D)\frac{\pi(\mathbf{x};\gamma^*)}{1-\pi(\mathbf{x};\gamma^*)}\right]\cdot\frac{E\left[(1-D)\frac{\pi(\mathbf{x}_i;\gamma^*)}{1-\pi(\mathbf{x};\gamma^*)}\mathbb{1}\{\Delta Y_t \leq \delta\}\right]}{E\left[(1-D)\frac{\pi(\mathbf{x};\gamma^*)}{1-\pi(\mathbf{x};\gamma^*)}\right]^2} + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}(\tilde{w}_0(D_i,\mathbf{x}_i;\hat{\gamma})\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&- n^{-1/2}\sum_{i=1}^{n}((\tilde{w}_0(D_i,\mathbf{x}_i;\hat{\gamma}) - 1)E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}((\tilde{w}_0(D_i,\mathbf{x}_i;\hat{\gamma})(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_{ti} \leq \delta\}]) + o_p(1)
\end{aligned}$$

where

$$\tilde{w}_0(D, \mathbf{x}; \hat{\gamma}) = \left. \frac{\pi(\mathbf{x}; \hat{\gamma})(1 - D)}{1 - \pi(\mathbf{x}; \hat{\gamma})} \right/ E\left[\frac{\pi(X; \gamma^*)(1 - D)}{1 - \pi(\mathbf{x}; \gamma^*)}\right]$$

Then, I do a second-order Taylor expansion around γ^*, so that

$$\begin{aligned}
&\sqrt{n}(C\hat{D}F^1 - CDF^1) \\
&= n^{-1/2} \sum_{i=1}^{n} w_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&+ (\hat{\gamma} - \gamma*)' \cdot n^{-1/2} \sum_{i=1}^{n} \dot{w}_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}] + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^{n} w_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&+ \sqrt{n}(\hat{\gamma} - \gamma*)' \cdot n^{-1} \sum_{i=1}^{n} \dot{w}_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}] + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^{n} w_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&+ n^{-1/2} \sum_{i=1}^{n} l_{\gamma*}(W_i)' \cdot E[\dot{w}_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}])] + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^{n} w_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&+ n^{-1/2} \sum_{i=1}^{n} l_{\gamma*}(W_i)' \cdot E[\alpha(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}])\dot{\pi}(X; \gamma*)] + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^{n} (w_0(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}]) \\
&+ l_{\gamma*}(W_i)' \cdot E[\alpha(D_i, \mathbf{x}_i; \gamma*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - E[w_0(D, \mathbf{x}; \gamma*)\mathbb{1}\{\Delta Y_t \leq \delta\}])\dot{\pi}(\mathbf{x}; \gamma*)]) + o_p(1)
\end{aligned}$$

where

$$\dot{w}(D, \mathbf{x}; \gamma) = \alpha(D, \mathbf{x}; \gamma)\dot{\pi}(\mathbf{x}; \gamma)$$

$$\alpha(D,\mathbf{x};\gamma) = \frac{1-D}{(1-\pi(\mathbf{x};\gamma))^2} \Bigg/ E\left[\frac{\pi(\mathbf{x};\gamma^*)(1-D)}{1-\pi(\mathbf{x};\gamma^*)}\right]$$

Observe that,

$$\begin{aligned}
\hat{CDF}^2 - CDF^2 &= n^{-1}\sum_{i,j=1}^{n}\left[\left(\frac{1-D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\frac{\pi(\mathbf{x}_i;\hat{\gamma})}{1-\hat{\pi}(\mathbf{x}_i;\hat{\gamma})}\left[\frac{(1-D_j)(\mathbb{1}\{\Delta\mu_{0t}(\mathbf{x}_i;\hat{\beta})+\Delta\hat{u}_{0tj}\})}{n_{1-D}}\right]\right]\right] - E\left[\left(\frac{1-D_i}{p}\frac{\tilde{p}(\mathbf{x}_i;\gamma)}{1-\pi(\mathbf{x}_i;\gamma)}\right)\tilde{P}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)\right] \\
&- n^{-1}\sum_{i,j=1}^{n}\left[\left(\frac{D_i}{\frac{\sum_{k=1}^{N} D_k}{n}}\left[\frac{(1-D_j)(\mathbb{1}\{\Delta\mu_{0t}(\mathbf{x}_i;\hat{\beta})+\Delta\hat{u}_{0tj}\})}{n_{1-D}}\right]\right]\right] - E\left[\left(\frac{D}{p}\right)\tilde{P}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)\right] \\
&= (\hat{CDF}^{21} - CDF^{21}) - (\hat{CDF}^{22} - CDF^{22})
\end{aligned}$$

Similarly note that,

$$\begin{aligned}
&\sqrt{n}(\hat{CDF}^{22} - CDF^{22}) \\
&= n^{-1/2}\sum_{i=1}^{n} w_1(D_i)\left(\tilde{P}(\Delta Y_{0ti}\le\delta|\mathbf{x}_i;\beta^*) - E[w_1(D)\tilde{P}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)]\right) \\
&+ \sqrt{n}(\hat{\beta}-\beta^*)'\cdot n^{-1}\sum_{i=1}^{n}(w_1(D_i)\dot{\tilde{P}}(\Delta Y_{0t}\le\delta|\mathbf{x}_i;\beta^*)) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n} w_1(D_i)\left(\tilde{P}(\Delta Y_{0ti}\le\delta|\mathbf{x}_i;\beta^*) - E[w_1(D)\tilde{P}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)]\right) \\
&+ n^{-1/2}\sum_{i=1}^{n} l(W_i;\beta^*)'E[w_1(D_i)\dot{\tilde{P}}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)] + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n}(w_1(D_i)\left(\tilde{P}(\Delta Y_{0ti}\le\delta|\mathbf{x}_i;\beta^*) - E[w_1(D)\tilde{P}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)]\right) \\
&+ l_{\beta^*}(W_i)'E[w_1(D_i)\dot{\tilde{P}}(\Delta Y_{0t}\le\delta|\mathbf{x};\beta^*)]) + o_p(1)
\end{aligned}$$

where $w_1(D) = \frac{D}{\frac{\sum_{k=1}^{n} D_k}{n}}$ Furthermore, note that

$$\begin{aligned}
&\sqrt{n}(\hat{CDF}^{21} - CDF^{21}) \\
&= n^{-1/2}\sum_{i=1}^{n}\tilde{w}_1(D_i,\mathbf{x}_i;\hat{\gamma})(\tilde{P}(\Delta Y_{0ti}\le\delta;|\mathbf{x}_i\hat{\beta}) - E[w_0(D,\mathbf{x}\ \gamma^*)\tilde{P}(\Delta Y_{0ti}\le\delta|\mathbf{x};\beta^*)] + o_p(1)
\end{aligned}$$

$$= n^{-1/2}\sum_{i=1}^{n} w_1(D_i, X_i;\gamma^*)(\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) - E[w_0(D,\mathbf{x};\gamma)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)]$$

$$+\ \sqrt{n}(\hat{\gamma}-\gamma^*)' \cdot E\left[\alpha(D,\mathbf{x};\gamma^*)\left(\tilde{P}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*) - E[w_0(D,\mathbf{x};\gamma^*)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)]\right)\dot{\pi}(\mathbf{x};\gamma^*)\right]$$

$$+\ \sqrt{n}(\hat{\beta}-\beta^*)' \cdot E[w_0(D,\mathbf{x};\gamma^*)\dot{\tilde{P}}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)] + o_p(1)$$

$$= n^{-1/2}\sum_{i=1}^{n}(w_0(D_i,\mathbf{x}_i;\gamma^*)(\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) - E[w_0(D,\mathbf{x};\gamma^*)P(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)])$$

$$+ l_{\gamma^*}(W_i)' \cdot E\left[\alpha(D,\mathbf{x};\gamma^*)\left(\tilde{P}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*) - E[w_0(D,\mathbf{x};\gamma^*)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)]\right)\dot{\pi}(\mathbf{x};\gamma^*)\right]$$

$$+ l_{\beta^*}(W_i)' \cdot E[w_0(D,\mathbf{x};\gamma^*)\dot{\tilde{P}}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)]) + o_p(1)$$

Then by combining all the asymptotic expansions, I obtain

$$\sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(\delta))$$

$$= n^{-1/2}\sum_{i=1}^{n}(w_0(D_i,\mathbf{x}_i;\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \le \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \le \delta\}])$$

$$+ l_{\gamma^*}(W_i)' \cdot E[\alpha(D_i,\mathbf{x}_i;\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \le \delta\} - E[w_0(D,\mathbf{x};\gamma^*)\mathbb{1}\{\Delta Y_t \le \delta\}])\dot{\pi}(\mathbf{x};\gamma^*)]$$

$$+ [w_1(D_i)\left(\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) - E[w_1(D)\tilde{P}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*)]\right)$$

$$+ l_{\beta^*}(W_i)' E[w_1(D_i)\dot{\tilde{P}}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*)]]$$

$$- [(w_0(D_i,\mathbf{x}_i;\gamma^*)(\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) - E[w_0(D,\mathbf{x};\gamma^*)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)])$$

$$+ l_{\gamma^*}(W_i)' \cdot E\left[\alpha(D,\mathbf{x};\gamma^*)\left(\tilde{P}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*) - E[w_0(D,\mathbf{x};\gamma^*)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)]\right)\dot{\pi}(\mathbf{x};\gamma^*)\right]$$

$$+ l_{\beta}(W_i)' \cdot E[w_0(D,\mathbf{x};\gamma^*)\dot{\tilde{P}}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)])] + o_p(1)$$

After simplification, I obtain,

$$\sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(\delta))$$

$$= n^{-1/2}\sum_{i=1}^{n}(w_0(D_i,\mathbf{x}_i;\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \le \delta\} - \tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) + E[w_0(\gamma)\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x};\beta^*)] - E[w_0\mathbb{1}\{\Delta Y_t \le \delta\}])$$

$$+ [w_1(D_i)\left(\tilde{P}(\Delta Y_{0ti} \le \delta|\mathbf{x}_i;\beta^*) - E[w_1\tilde{P}(\Delta Y_{0t} \le \delta|\mathbf{x};\beta^*)]\right)$$

$$+ l_{\beta^*}(W_i)' E[(\dot{\tilde{P}}(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*))(w_1 - w_0)]$$

$$+ l_{\gamma^*}(W_i)' E[\alpha(\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} + \tilde{P}(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*)) - E[w_0(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - \tilde{P}(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*)])\dot{\pi}(\gamma^*)]] + o_p(1)$$

Now, suppose that the propensity score and the CDF of $\Delta Y_{0ti}|X$ are correctly specified.

Note that $l_{\beta^*}(W_i)' E[(\dot{P}(\Delta Y_{0ti} \leq \delta;\beta^*))(w_1 - w_0)] = 0$,

$l_{\gamma^*}(W_i)' E[\alpha(\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} + P(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*)) - E[w_0(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - P(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*)])\dot{\pi}(\gamma^*)]] =$

0, and $E[w_0(\gamma^*)P(\Delta Y_{0ti} \leq \delta|\mathbf{x};\beta^*)] - E[w_0\mathbb{1}\{\Delta Y_t \leq \delta\}] = 0$. Then,

$$\sqrt{n}(\hat{F}_{\Delta Y_{0t}|D=1}(\delta) - F_{\Delta Y_{0t}|D=1}(\delta))$$
$$= n^{-1/2}\sum_{i=1}^{n}\Big[w_0(D_i,\mathbf{x}_i;\gamma^*)(\mathbb{1}\{\Delta Y_{ti} \leq \delta\} - (w_0(D_i,\mathbf{x}_i;\gamma^*) - w_1(D_i,\mathbf{x}_i;\gamma^*))P(\Delta Y_{0ti} \leq \delta|\mathbf{x}_i;\beta^*) - w_1(D_i)F_{\Delta Y_{0t}|D=1}(\delta)\Big]$$
$$+ o_p(1)$$
$$= n^{-1/2}\sum_{i=1}^{n}\psi(D_i,\mathbf{x}_i,Y_{0i},Y_{1i}) + o_p(1)$$

□

Proof of Proposition 1:

Proof. The result follows from Theorem 3 and the functional central limit theorem for empirical distribution functions. □

Proof of Proposition 2

Proof. The result follows by Proposition 1, Lemma B.4 in Callaway and Li (2019), and similar arguments used to establish Proposition 4 in Callaway and Li (2019). □

Proof of Theorem 4: The result follows from Proposition 2 and Lemma 3.9.23(ii) in Vaart and Wellner (1996).

Outline of proof of Proposition 3 Consider the asymptotic expansion in the nonparametric case of Theorem 5. Consider each term, assuming that the bootstrap estimate of each nuisance function converges to the estimate of each function based upon the unweighted data. I will consider the first term in the expansion It remains to show that each term is $o_p(n^{1/2})$. Each term outside of the first term converges converges either due to Rothe and Firpo (2019) or due to the fact that $sup_{\mathbf{x}\in\mathcal{X}}|\hat{\pi}^*(\mathbf{x}_i) - \hat{\pi}(\mathbf{x}_i)| = o_p(n^{1/4})$. For the first term, let $\mathbb{G}_n(f_1) = n^{1/2}(\mathbb{P}_n - \mathbb{P})f_0(D, \Delta Y_t, \mathbf{x}^*)$, where $\mathcal{P}_n$ is the empirical measure, $\mathcal{P}$ is the expectation, and

$$f_1^*(D, \Delta Y_t, \mathbf{x}) = \frac{(1-D)(\mathbb{1}_{\Delta Y - t \leq y} - \hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}^*))}{p^*(1-\hat{\pi}(\mathbf{x}^*))^2}(\hat{\pi}^*(\mathbf{x}_i^*) - \hat{\pi}(\mathbf{x}_i^*))$$

Since $sup_{x\in\mathcal{X}}|\hat{\pi}^*(\mathbf{x}) - \hat{\pi}(\mathbf{x})| \lesssim o_p((k_n^{-a/d}) = o_p(1)$ by Theorem 3.2 in Chen (2007) and the previous proof, then define $\mathcal{F} = \{f_1 : ||\hat{\pi}^*(\mathbf{x}) - \hat{\pi}(\mathbf{x})||_\infty \leq \delta_{1n}, ||\hat{P}^*(\mathbf{x}) - \hat{P}(\mathbf{x})||_\infty \leq \delta_{2n}\}$, where $\delta_{1n} = C(k_n^{-a/d})$ for some $C > 0$, and $\delta_{2n} = K^{-r}$, where $r > 1/2$. Let $\mathcal{G}_1 = \{\pi \in \Lambda_c^p(\mathcal{X}) : ||\pi - \hat{\pi}||_\infty \leq \delta_{1n}$ and $\mathcal{G}_2 = \{P \in \mathcal{M} : ||P - \hat{P}||_\infty \leq \delta_{2n}\}$, where $\mathcal{M}$ denotes a Hölder space containing an estimate of $P(\Delta Y_{0t} \leq y|\mathbf{x})$, such as the kernel estimator mentioned in this text, with smoothing parameter a_1 such that $d/a_1 \geq 2$. Note the following

$$\frac{\sum_{i=1}^n \psi_i^1(\hat{\pi}^*(x^*) - \hat{\pi}(x^*))}{n^{1/2}} \leq sup_{f_1\in\mathcal{F}}\mathbb{G}_n(f_1) + n^{1/2}sup_{f_1\in\mathcal{F}}\mathbb{P}f_1$$

I will consider the second term first. Since p and $\pi(x)$ are bounded away from 0, then

$$\begin{aligned}
n^{1/2}sup_{f_0\in\mathcal{F}}\mathbb{P}f_1 &= n^{-1/2}sup_{\pi\in\mathcal{G}1, P\in\mathcal{G}_2}E\left[\left(\frac{1-D}{p^*}\frac{1}{(1-\hat{\pi}(\mathbf{x}^*))^2}\right)[\mathbb{1}\{\Delta Y_t \leq y\} - P(\Delta Y_{0t} \leq y|\mathbf{x}_i)](\hat{\pi}^*(\mathbf{x}^*) - \hat{\pi}(\mathbf{x}^*)\right] \\
&\lesssim n^{1/2}sup_{x\in\mathcal{X}}|\hat{\pi}^*(\mathbf{x}^*) - \hat{\pi}(\mathbf{x}^*)|sup_{x\in\mathcal{X}}|\hat{P}^*(\mathbf{x}^*) - \hat{P}(\mathbf{x}^*)| \\
&\lesssim o_p(1)
\end{aligned}$$

where the last line follows from Theorem 3. Now, I will consider the term $sup_{f_1\in\mathcal{F}}\mathbb{G}_n(f_1)$. Let $F_1 := \delta_n = C(k_n^{-a/d})$, so $||F||_{P,2} \lesssim \delta_n$. Let $\mathcal{F}_1 = \{f_1 : ||\hat{\pi}^* - \pi||_\infty \leq C, ||\hat{P}^* - P|| \leq 1\}$. Define $\mathcal{F}_{10} = \{\pi \in \Lambda_c^p(\mathcal{X}) + \hat{\pi}^* : ||\pi||_{p,2} \leq C\}$ and $\mathcal{F}_{20} = \{P \in \mathcal{M} + \hat{P}^* : ||P||_{p,2} \leq 1\}$. Then,

$$logN_{[]}(\epsilon, \mathcal{F}, L_2(p)) \lesssim logN_{[]}(\epsilon/\delta_{2n}, \mathcal{F}_0, L_2(p)$$

$$\lesssim logN_{[]}(\epsilon/\delta_{2n}, \mathcal{F}_{10}, L_2(p) + logN_{[]}(\epsilon/\delta_{2n}, \mathcal{F}_{20}, L_2(p)$$
$$\lesssim logN_{[]}(\epsilon/\delta_{2n}, \Lambda_c^p(\mathcal{X}), L_2(p) + logN_{[]}(\epsilon/\delta_{2n}, \mathcal{M}, L_2(p)$$
$$\lesssim (\delta_n/\epsilon)^{d/a} + (\delta_n/\epsilon)^{d/a_1}$$

This is sufficient to demonstrate that the bracketing integral converge of $J_{[]}(\|F_1\|_{p,2}, \mathcal{F}, L_2(p))$ converges. Then $\frac{\sum_{i=1}^n \psi_i^1(\hat{\pi^*}(x^*) - \hat{\pi}(x^*))}{n^{1/2}} = o_p(1)$.

Now, note that

$$E\left[\left(-\frac{1-D_i}{p^*}\frac{\hat{\pi}^*(\mathbf{x}^*{}_i)}{1-\hat{\pi}^*(\mathbf{x}^*{}_i)} + \frac{D_i}{p^*}\right)\left(\hat{P}^*(\Delta Y_{0t} \leq y|\mathbf{x}^*) - \hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}^*)\right)\right]$$
$$\lesssim E\left[(\hat{\pi}^*(\mathbf{x}) - D_i)\left(\hat{P}^*(\Delta Y_{0t} \leq y|\mathbf{x}^*) - \hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}^*)\right)\right]$$
$$\lesssim E\left[(|\hat{\pi}^*(\mathbf{x}^*) - \hat{\pi}(\mathbf{x}^*)| + |\hat{\pi}(\mathbf{x}^*) - \pi(\mathbf{x})|)\left(\hat{P}^*(\Delta Y_{0t} \leq y|\mathbf{x}^*) - \hat{P}(\Delta Y_{0t} \leq y|\mathbf{x}^*)\right)\right]$$

The result then follows by the same steps used to show that $\frac{\sum_{i=1}^n \psi_i^1(\hat{\pi}^*(x^*) - \hat{\pi}(x^*))}{n^{1/2}}$ is $o_p(1)$. Then the main result follows by Theorem 3.6.1 in Vaart and Wellner (1996).

Proof of Theorem 5:

Proof. The result follows by Proposition 3, Lemma 3.9.23(ii), and Theorem 3.9.11 in Vaart and Wellner (1996). □

C Tables and Figures

Table 1: $QTT(0.5)$ Estimates

	$\widehat{QTTdr}$	$\widehat{QTTpro}$	$\widehat{QTTcdf}$	$\widehat{QTTcl}$	$\widehat{QTTno}$
$\tau = 0.5$					
QTT	-2.670	-2.670	-2.670	-2.670	-2.670
Estimate	-2.679	-2.688	-2.684	-2.517	-2.685
se	0.140	0.143	0.150	1.504	0.150
RMSE	0.141	0.143	0.150	1.480	0.150
N	1000	1000	1000	1000	1000
T	3	3	3	3	3

Figure 1: Average Absolute Bias

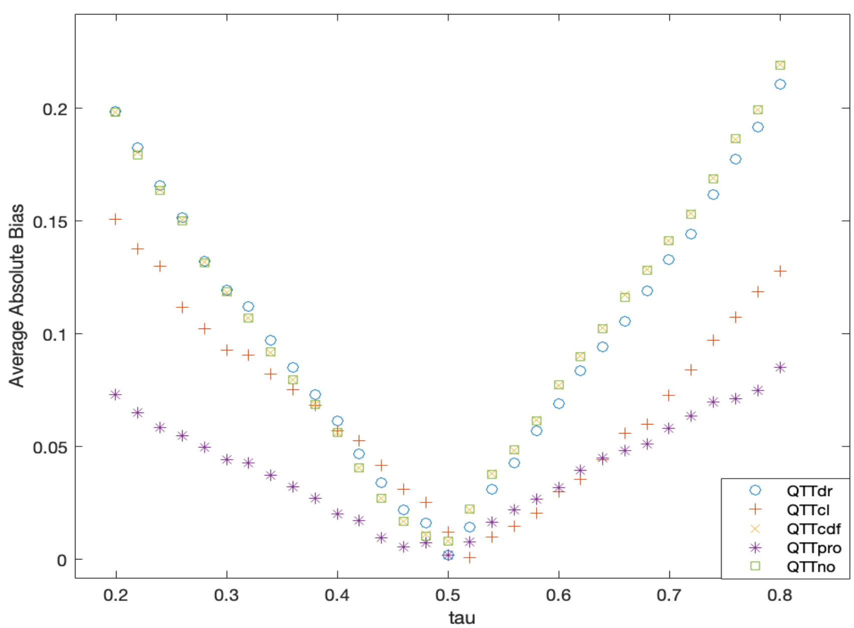

Figure 1: $QTTdr$ represents the estimates when both nuisance functions are correctly specified. $QTTpro$ represents the estimates when the propensity score is correctly specified, but the conditional cdf is incorrectly specified. $QTTcdf$ represents the analogous estimates to $QTTpro$ when the conditional cdf is correctly specified. $QTTcl$ represents the Callaway and Li estimates. $QTTno$ represents when neither of the nuisance functions are correctly specified.

Figure 2: RMSE

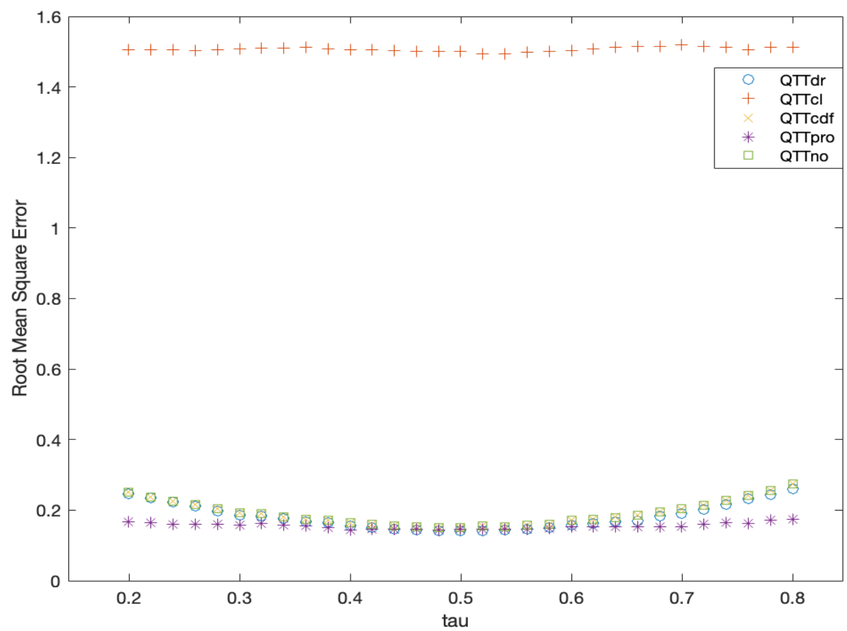

Figure 2: $QTTdr$ represents the estimates when both nuisance functions are correctly specified. $QTTpro$ represents the estimates when the propensity score is correctly specified, but the conditional cdf is incorrectly specified. $QTTcdf$ represents the analogous estimates to $QTTpro$ when the conditional cdf is correctly specified. $QTTcl$ represents the Callaway and Li estimates. $QTTno$ represents when neither of the nuisance functions are correctly specified.

Figure 3: QTT Unemployment Estimates

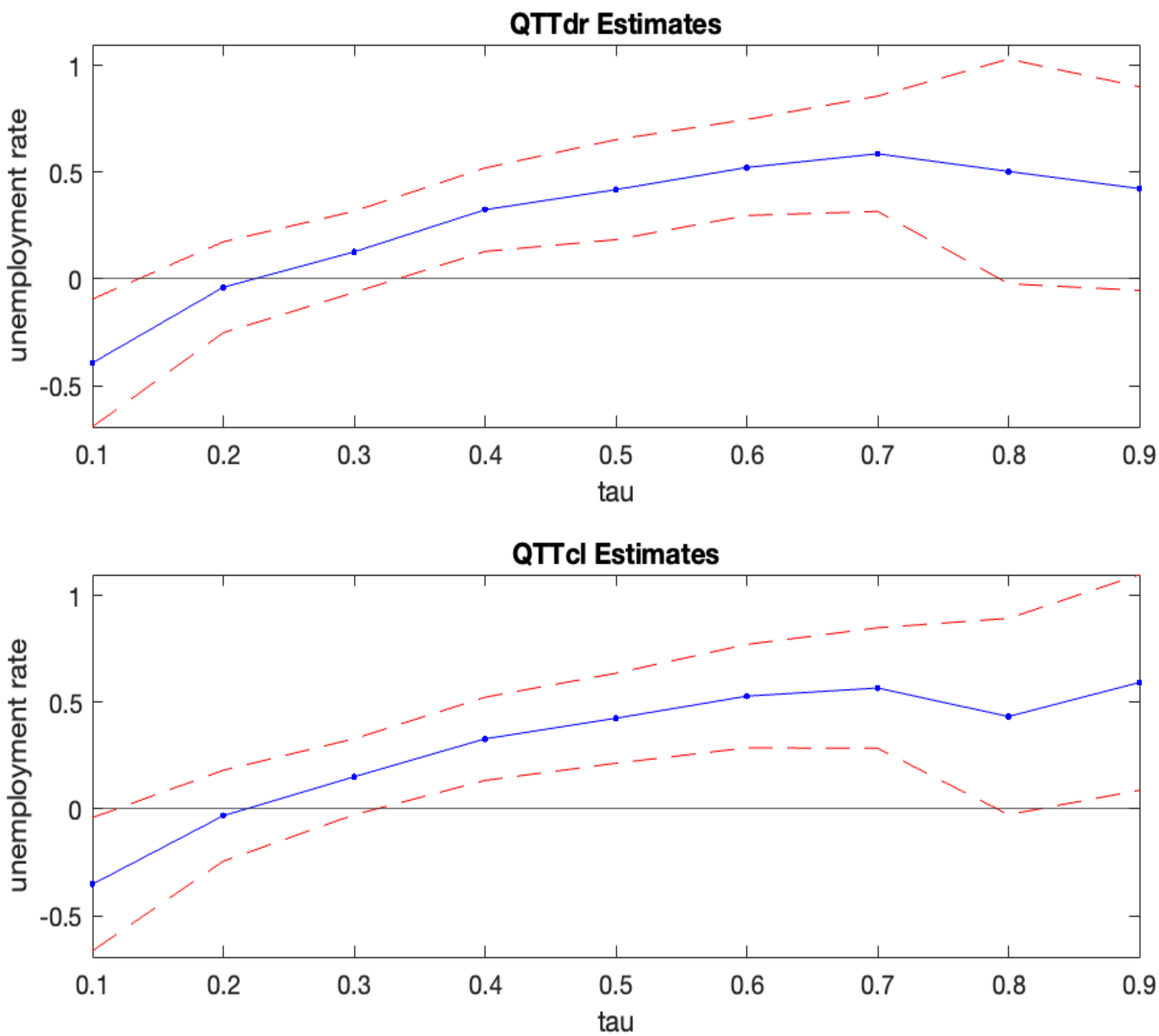

Figure 3: The top panel represents estimates of the $QTT(\tau)$ and their confidence intervals using my doubly-robust estimator. The bottom panel represents the estimates of the $QTT(\tau)$ using the Callaway and Li estimator. The blue line represents the curve of point estimates. The red lines represent the 95% confidence bonds.